

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 16:09:48 ; Search time 24.3 Seconds

(without alignments) updates/sec
2393.655 Million cell

Title: US-09-468-147-91

Perfect score: 9013
Sequence: 1 PGITTAIEQALAAANSALA.....FTETIKVEDLTNSIIOIVE 1698

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7459	82.8	1693	1	P33424 hepatitis e
2	7448	82.6	1693	1	POLN_HEVPA
3	7433	82.5	1691	1	POLN_HEVME
4	7433	82.5	1693	1	POLN_HEVME
5	406.5	4.5	1648	1	RRPO_CGMS
6	374.5	4.2	2205	1	POLN_RUBVT
7	351.5	3.9	1612	1	RRPO_ORSVS
8	349	3.9	1844	1	POLR_TYV
9	346	3.8	1844	1	POLR_TYV
10	339.5	3.8	1612	1	RRPO_ORSYC
11	322.5	3.6	1616	1	RRPO_TOML
12	321.5	3.6	1616	1	RRPO_TOML
13	320.5	3.6	1616	1	RRPO_TOML
14	320	3.6	1597	1	RRPO_CRMV
15	319.5	3.5	1616	1	RRPO_TOMB
16	316.5	3.5	1616	1	RRPO_TOMK2
17	313	3.5	1601	1	RRPO_TYCV
18	312.5	3.5	1612	1	RRPO_PMYV
19	312.5	3.5	1612	1	RRPO_PMYV
20	311	3.5	1748	1	POLR_ELV
21	310	3.4	1844	1	POLR_TYV
22	305.5	3.4	1776	1	POLR_OYV
23	301.5	3.3	1616	1	RRPO_TMYOM
24	299.5	3.3	1616	1	RRPO_TMYB
25	295.5	3.3	1616	1	RRPO_TMYV
26	293.5	3.3	1616	1	RRPO_TMYV
27	293	3.3	1874	1	POLR_KYV
28	291.5	3.2	1616	1	RRPO_TMYV
29	287.5	3.2	1839	1	POLR_EPV
30	261.5	2.9	1609	1	RRPO_TMYV
31	254	2.8	1629	1	RRPO_SHV
32	243.5	2.7	1707	1	194K_TRVSV
33	243.5	2.7	1884	1	RRPO_ACLS

34	242.5	2.7	1885	1	RRPO_ACLS	P54891 apple chlor
35	218	2.4	2512	1	POLN_SINDV	P03317 sindbis vir
36	211	2.3	2514	1	POLN_SINDO	P27283 sindbis vir
37	207	2.3	2514	1	POLN_ONNNG	P13886 o'nyong-nyo
38	201.5	2.2	1718	1	RRPO_SHVX	P04575 shalilot vir
39	200	2.2	2431	1	POLN_SFV	P08411 semliot for
40	194	2.2	1967	1	RRPO_PVMR	P17965 potarto viru
41	194	2.2	2485	1	POLN_EEVV3	P36327 venezuelan
42	184.5	2.0	3175	1	RRPO_EAV	P18611 equine arte
43	177.5	2.0	2492	1	POLN_EEVT	P27282 venezuelan
44	176	2.0	1385	1	RRPO_PIAMV	P07518 plantago as
45	174.5	1.9	2492	1	POLN_EEVP	P36328 venezuelan

ALIGNMENTS

RESULT 1
ID POLN_HEVPA STANDARD: PRT: 1693 AA.

AC P33424:
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
(EC 2.7.7.48); HELICASE].
OS Hepatitis E virus (strain Pakistan) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_Taxid=33774;
RN [1]

SEQUENCE FROM N.A.
RA MEDLINE=92115700; PubMed=1731327;
RA Tsarev S.A., Emerson S.O., Reyes G.R., Tsareva T.S., Legters L.J.,
RA Malik I.A., Iqbal M., Purcell R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
CC -!- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).

CC EMBL: M80581; AAA5725.1; -.
DR MEROPS: C41.001; -.
DR InterPro: IPR000606; -.
DR InterPro: IPR002588; -.
DR InterPro: IPR002589; -.
DR Pfam: PF01461; DUF27; 1.
DR Pfam: PF01443; Viral_helicase; 1.
DR Pfam: PF01660; Methyltransf; 1.
KW Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;
KW ATP-binding; 975
FT NP_BIND
SQ SEQUENCE 1693 AA; 185149 MW; 5A0F03FB1FB99E8 CRC64;

Query Match 82.8%; Score 7459; DB 1; Length 1693;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1401; Conservative 109; Mismatches 160; Indels 42; Gaps 7;

QY	1	PGITTAIEQALAAANSALANVAVPFLSRVOTELLINIMAPROLVFPEVLNHPRIOR	60
DB	10	PGITTAIEQALAAANSALANVAVPFLSRVOTELLINIMAPROLVFPEVLNHPRIOR	69
QY	61	VHNELEIYCARAGCLEVGAHPSINDPNVLAHRCFLRPGVDVQRYSAFTRCPPAAN	120
DB	70	VHNELEIYCARAGCLEVGAHPSINDPNVLAHRCFLRPGVDVQRYSAFTRCPPAAN	129

```

QY 121 CRRSALRGLPDRATYCFDFSCRCFAETGVALYSLHMDPADVABAMAHGXTRLYAA 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 130 CRRSALRGLPDRATYCFDFSCRCFAETGVALYSLHMDSPDVAEMRPHGXTRLYAA 189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 181 LHLPEVLLPRTYHTSYLLIHGDDRAVYTYEGDTSAGYVHDSILRAVTRTKYGDH 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 LHLPEVLLPRTYHTSYLLIHGDDRAVYTYEGDTSAGYVHDSILRAVTRTKYGDH 249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 241 PLVIERAIGCHFVLLTFAPEPSMPYVYPRSTEVYASISPGSGSPSLFSPSASTK 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 250 PLVIERAIGCHFVLLTFAPEPSMPYVYPRSTEVYASISPGSGSPSLFSPSASTK 309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 301 STEHAVVHIMDLMLFGATLDQAFCSSRLMTYLRGISYKYVYGVALVANGMASDAL 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 310 STEHAVVHIMDLMLFGATLDQAFCSSRLMTYLRGISYKYVYGVALVANGMASDAL 369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 361 TAYTAATVITICHOXYLRTQATISGMRLGVEHAKFTTRYSMLFPEKSGDYIPGRQLO 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 370 TAYTAATVITICHOXYLRTQATISGMRLGVEHAKFTTRYSMLFPEKSGDYIPGRQLO 429
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 421 FYAOCRRMLSGFHLDPRLVLFDESVPCRCRTFLKAYAGKFCFPMRLQDCTCELEPAE 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 430 FYAOCRRMLSGFHLDPRLVLFDESVPCRCRTFLKAYAGKFCFPMRLQDCTCELEPAE 489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 481 GLVGDHNDNAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYALVNDIARASRLT 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 490 GLVGDHNDNAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYALVNDIARASRLT 549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 541 ATVELVASPDLERCTVYLGKNTFRTYVVDGAHLBANGPEQVLSFSDASROSGASLSLT 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 550 ATVELVASPDLERCTVYLGKNTFRTYVVDGAHLBANGPEQVLSFSDASROSGASLSLT 609
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 601 YELTPAGLQVRISSNGLDCTATPPGAPASAPGEVAAFSCALRYNRTQHSLSLGLW 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 610 YELTPAGLQVRISSNGLDCTATPPGAPASAPGEVAAFSCALRYNRTQHSLSLGLW 669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 661 LHPGGLIGTPPSPGHTWESANPFCGEGTYLTKWS-----TSGFSSDFSPPE 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 670 LHPGGLIGTPPSPGHTWESANPFCGEGTYLTKWS-----TSGFSSDFSPPE 724
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 710 AAAPAMATGCLPHSTPVPVSDIWLPPSEEFQVDAAP-PPAPDPA-GLGPGVVLPPP 767
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 725 PSLPSPRAATP-----TP-----AAPLPPPADPSPPTLSAPRAGEBAP 761
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 768 PPVYKPSIP-PPSRNRLLYTPYDGAQKVVAGSLFESDCDMLVNASVNGHRCGGLCHAF 826
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 762 GATARAATHTHQRHRRRLFTYPDGSKVAFGSLFESTCTMLVNASVNDHRPGGLCHAF 821
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 827 YORPEPAFYPTERTMREGLAAYTLTPRPIIHAVAPDYRVEDONKRLDEAYRETCSRGT 886
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 822 YORPEPAFYPTERTMREGLAAYTLTPRPIIHAVAPDYRVEDONKRLDEAYRETCSRGT 881
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 887 AYPLGSGIYOVPVLSFDMERNHRPDELYLPELAAWFEARPPCPPLTLTEDVART 946
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 882 AYPLGSGIYOVPVLSFDMERNHRPDELYLPELAAWFEARPPCPPLTLTEDVART 941
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 947 ANLALEIDAATEVGRACAGCTISPGIYHOYTAGVPGSGKRSIQGCDVYVVPTRRLR 1006
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 942 ANLALEIDAATEVGRACAGCTISPGIYHOYTAGVPGSGKRSIQGCDVYVVPTRRLR 1001
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1007 NSNRRCGAFTPHRTAARVITIGRRVVIDEAPSLPHLLLHMORASSVHLLGDPNQIPAI 1066
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1002 NSNRRCGAFTPHRTAARVITIGRRVVIDEAPSLPHLLLHMORASSVHLLGDPNQIPAI 1061
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1067 DEFHAGLVPAIRBELAPTSMMXYTHRCPADVCELIRGAYKIQTSVLSLFENEPAG 1126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1062 DEFHAGLVPAIRBELAPTSMMXYTHRCPADVCELIRGAYKIQTSVLSLFENEPAG 1121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1127 OKLVXTQAAANPAGATVVEAGATETTTIINTADARGIIOSSRAHAVALTRHTEKC 1186
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1122 OKLVXTQAAANPAGATVVEAGATETTTIINTADARGIIOSSRAHAVALTRHTEKC 1181
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1187 VIIDAPGLLEVEGISDVIVNNFLAGSEVGHRSPIIRGNPDONLGTLOAFPSCQISA 1246
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db 1182 VIIDAPGLLEVEGISDVIVNNFLAGSEVGHRSPIIRGNPDONLGTLOAFPSCQISA 1241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1247 YHOLAEELGHRPAPVAVLPPCPPELBOGILYMPDELVSQSVYFEELDIYHCHMAAPSO 1306
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1242 YHOLAEELGHRPAPVAVLPPCPPELBOGILYMPDELVSQSVYFEELDIYHCHMAAPSO 1301
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1307 RKAVALSTLVGRYGRRTKLYEASHDSRESLARIPRTIGPVQATTCBELYELVAMVERGOD 1366
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1302 RKAVALSTLVGRYGRRTKLYEASHDSRESLARIPRTIGPVQATTCBELYELVAMVERGOD 1361
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1367 GSAVLELDLCNRDYSRTTFQKXNKRTTGETTAHGVVGGISAMSTFCALRPPMRRAI 1426
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1362 GSAVLELDLCNRDYSRTTFQKXNKRTTGETTAHGVVGGISAMSTFCALRPPMRRAI 1421
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1427 EKETALLPPIEYGDYVESVFAAASGAGSCVAFENDSEFPSTONNLSIGLEYVME 1486
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1422 EKETALLPPIEYGDYVESVFAAASGAGSCVAFENDSEFPSTONNLSIGLEYVME 1481
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1487 ECGMPQWLIRLYHLVRSAMILLQAPKESLKGFWKXSGEPGTLMTVNMMAIYAHCEYFR 1546
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1482 ECGMPQWLIRLYHLVRSAMILLQAPKESLKGFWKXSGEPGTLMTVNMMAIYAHCEYFR 1541
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1547 DEFVAARFGDSDVYLCSDYQSRNAALINGCGIKLVDRPRTGLYAGVVAAPGLTLPD 1606
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1542 DEFVAARFGDSDVYLCSDYQSRNAALINGCGIKLVDRPRTGLYAGVVAAPGLTLPD 1601
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1607 VVRPAGRLSEKMMGPGPERAEQRLAICDFLRGLTNNYAQVCVDVYASVYGVSPGLVNL 1666
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1602 VVRPAGRLSEKMMGPGPERAEQRLAICDFLRGLTNNYAQVCVDVYASVYGVSPGLVNL 1661
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1667 GMLQTIADGKAHFTETIKPVLDLNTSIIOVE 1698
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1662 GMLQTIADGKAHFTETIKPVLDLNTSIIOVE 1693
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

RESULT 2
 ID POLN_HEVBU STANDARD; PRT; 1693 AA.
 AC P29324;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DE NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48); HELICASE].
 OS Hepatitis E virus (strain Burma) (HEV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage.
 ON NCBI_TaxID=31767;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92024067; PubMed=1926770;
 RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
 RA Fry K.E., Reyes G.R.;
 RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
 RT full-length viral genome.";
 RL Virology 185:120-131(1991).
 CC -
 CC MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
 CC ENTERICALLY TRANSMITTED NON-A; NON-B HEPATITIS (ET-NANBH).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch.)
 CC
 CC EMBL: M73218; AAA45734.1; -
 CC PIR: A40778; MNMME.
 CC MEROPS: C41.001; -
 CC InterPro: IPR000606; -
 CC InterPro: IPR002588; -
 CC InterPro: IPR002589; -

DR Pfam: PF01661: DUF27.1.
 DR Pfam: PF01443: Viral_helicase1.1.
 DR Pfam: PF01660: Methyltransferase.1.
 KW Polyprotein; transferase; RNA-directed RNA polymerase; Helicase;
 FT ATP-binding. 975
 NP_BIND 982 ATP (POTENTIAL).
 SQ SEQUENCE 1693 AA; 185191 MW; 2F355E469ED219B CRC64;

Query Match 82.6%; Score 7448; DB 1: Length 1693;
 Best Local Similarity 81.8%; Pred. No. 0;
 Matches 196; Conservative 110; Mismatches 169; Indels 32; Gaps 6;

QY 1 PGGTTTAEQAALAAANSALANAVVAPFLSRVOTELLINLMQPROLYFREVLMNPIQR 60
 DB 10 PGGTTTAEQAALAAANSALANAVVAPFLSRVOTELLINLMQPROLYFREVLMNPIQR 69
 QY 61 VHNLEQYCRARAGCLGVGAHPSRINDPNVLRHRCFLRPVGRDVRWYSAFTGPAAN 120
 DB 70 VHNLEQYCRARAGCLGVGAHPSRINDPNVLRHRCFLRPVGRDVRWYSAFTGPAAN 129
 QY 121 CRRSALRGLPPADRYTCFDFGSCFAETGVALYSLHDLMPADVAEAMARHGXTRLYAA 180
 DB 130 CRRSALRGLPPADRYTCFDFGSCFAETGVALYSLHDLMPADVAEAMARHGXTRLYAA 189
 QY 181 LHLPEVLLPPTGYTHTSTLLHDSRAVYTYEGDTSAGYNHDVSLTRAMRTTKVGDH 240
 DB 190 LHLPEVLLPPTGYTHTSTLLHDSRAVYTYEGDTSAGYNHDVSLTRAMRTTKVGDH 249
 QY 241 PLVIEVRAIGCFHVLTLAAREPSPMPVYPRSTEVVRSIFGSGSPSLPSCSTK 300
 DB 250 PLVIEVRAIGCFHVLTLAAREPSPMPVYPRSTEVVRSIFGSGSPSLPSCSTK 309
 QY 301 STPHAVVHIWDLMLFGATLDDQAFCCSRLMTYLRGISYKVTGALVANEGMNASDAL 360
 DB 310 STPHAVVHIWDLMLFGATLDDQAFCCSRLMTYLRGISYKVTGALVANEGMNASDAL 369
 QY 361 TXXITAAVLTICHOVLRTOALSCKMRRLGVEHAOKFETRLYSWLEKSGROVTPRQ 420
 DB 370 TXXITAAVLTICHOVLRTOALSCKMRRLGVEHAOKFETRLYSWLEKSGROVTPRQ 429
 QY 421 FYAOCRRMLISAGFHLDPVLYPDESVPKCRFLKRVAGKFCFCFMMWLGCECTLEPAR 480
 DB 430 FYAOCRRMLISAGFHLDPVLYPDESVPKCRFLKRVAGKFCFCFMMWLGCECTLEPAR 489
 QY 481 GLVGDHGHNEAESEVDPAEPALHDVSGTYAVHGHQLEALYRALNVPDIIAARSLT 540
 DB 490 GLVGDHGHNEAESEVDPAEPALHDVSGTYAVHGHQLEALYRALNVPDIIAARSLT 549
 QY 541 ATVELVADDRLECRVTLGKTKPRTVVGDAGHLEANGRPQVYLSFASRQSMGASHST 600
 DB 550 ATVELVADDRLECRVTLGKTKPRTVVGDAGHLEANGRPQVYLSFASRQSMGASHST 609
 QY 601 YELTPAGLOVRISNGLDCTATFPFGAPSAAGVEAALFALRYNRFORSLTGLW 660
 DB 610 YASASAGLEVRVAAGLDRAVFAPEVSPRSAPGEVATACSAIYRRNREQRSLIGNLM 669
 QY 661 LHPGLIGTFPPSPGHIMESANPCGEGTLTYRTWS-TSGESSDPS-PEAAPA 714
 DB 670 FHPGGLIGTFPPSPGHIMESANPCGEGTLTYRTWS-EVDAVVASPARDLGFMSSEPS 729
 QY 715 MATPGLPSTPVSVDIWLPPPESEFOYDAVY-PPAPDAPL-GPVYLTTPPPPVH 772
 DB 730 RAATPPL-AAAAAAAAAAAAAAAAAALPPAPPPPPAPALAEASAGATAG 766
 QY 773 KPSIP-PPSKRRLLTYPDGAKVYASLFESDCDWLVNANSNGHPRGGGLCHAFYQRP 831
 DB 767 APATHTQTAHRRLFTTYPDGSKVFAESLFESCTWLVNANSNDHPRGGGLCHAFYQRP 826
 QY 832 EAFYPTFETIRBGLAATYTLPRPIIAVADVAVVEONPKLEAAVETCSRGCTAAYPLL 891
 DB 827 ASFDAASFMRDGAAYTTLPRPIIAVADVAVVEONPKLEAAVETCSRGCTAAYPLL 886

QY 892 GSGIYQVPSLSTDAWERNHRCDELYLTPAANFEAKPKAPVLTITTEDTARTANL 951
 DB 887 GSGIYQVPSLSTDAWERNHRCDELYLTPAANFEAKPKAPVLTITTEDTARTANL 946
 QY 952 EIDPAETEVRACAGCTISGIYHYOFTAGVPGSGKRSIQGQDVVVVPTRELRSMR 1011
 DB 947 EIDPAETEVRACAGCTISGIYHYOFTAGVPGSGKRSIQGQDVVVVPTRELRSMR 1006
 QY 1012 RGFAPFTPTAAVITGRVYIDEAPSLPRLILLHMORASSVHLGDPNQIPADFEHA 1071
 DB 1007 RGFAPFTPTAAVITGRVYIDEAPSLPRLILLHMORASSVHLGDPNQIPADFEHA 1066
 QY 1072 GLVPAIRPELAPTSMMXVYHRCRADYCELLRGAYPKIOTTSRLSRSLFNNEAIGOKLV 1131
 DB 1067 GLVPAIRPELAPTSMMXVYHRCRADYCELLRGAYPKIOTTSRLSRSLFNNEAIGOKLV 1126
 QY 1132 TOAKANPAGAITVHEAOGATFETTLITADARGLIOSSRAHAYALTRHTEKCYILDA 1191
 DB 1127 TOAKANPAGAITVHEAOGATFETTLITADARGLIOSSRAHAYALTRHTEKCYILDA 1186
 QY 1192 PGLIREVIGISDVIVNFFLAGSEVGHXRPVYIPRGNDONLGLQAFPPSCQISATVHQLA 1251
 DB 1187 PGLIREVIGISDVIVNFFLAGSEVGHXRPVYIPRGNDONLGLQAFPPSCQISATVHQLA 1246
 QY 1252 EELGHRPAPVAAVLPCCPELEGLLYMPOELTVSDSVLVEFLTDIYHGMARSPORKAVL 1311
 DB 1247 EELGHRPAPVAAVLPCCPELEGLLYMPOELTVSDSVLVEFLTDIYHGMARSPORKAVL 1306
 QY 1312 STLVRGRTKLYLEAASDVRESLARFTPTGPOATTCCELYLEAVEWEEKODGSAYL 1371
 DB 1307 STLVRGRTKLYLEAASDVRESLARFTPTGPOATTCCELYLEAVEWEEKODGSAYL 1366
 QY 1372 ELDLCNRDYSRTTFPOKCKFTTGETTAHGVKGGISAMSKTFCLFQWPAITKELL 1431
 DB 1367 ELDLCNRDYSRTTFPOKCKFTTGETTAHGVKGGISAMSKTFCLFQWPAITKELL 1426
 QY 1432 ALLPPIFPGDAVEESVFAAASVAGSCWFEVDFEFPSTONFSLGECVAMECGMP 1491
 DB 1427 ALLPPIFPGDAVEESVFAAASVAGSCWFEVDFEFPSTONFSLGECVAMECGMP 1486
 QY 1492 QWLIRLYHLVRSAMILQAPRESLKGFKHSGEGTLNMTVMNMAIIACVFEFRFVA 1551
 DB 1487 QWLIRLYHLVRSAMILQAPRESLKGFKHSGEGTLNMTVMNMAIIACVFEFRFVA 1546
 QY 1552 AFKGDSSVLCSDYROSRNAALIAACGKLKVDYPRIGLYAGVVAAPGLTLPDYVRR 1611
 DB 1547 AFKGDSSVLCSDYROSRNAALIAACGKLKVDYPRIGLYAGVVAAPGLTLPDYVRR 1606
 QY 1612 GRLEKKNMGPPERABOLRLAVCDPLRGLTNVQVCVDVYSRVYVSGPLVYHNLIGMLQ 1671
 DB 1607 GRLEKKNMGPPERABOLRLAVCDPLRGLTNVQVCVDVYSRVYVSGPLVYHNLIGMLQ 1666
 QY 1672 IADKRAHFTETIKPVLDTNLSIIQRYE 1698
 DB 1667 VADKRAHFTESVAPVLDITNSICRYE 1693

RESULT 3
 AC 003495;
 ID POLN_HEVME STANDARD; PRT; 1691 AA.
 DT 01-OCT-1993 (rel. 27, Created)
 DT 01-DEC-1998 (rel. 37, Last sequence update)
 DE NON-STRUCTURAL POLYPROTEIN [CONTAINS: RNA-DIRECTED RNA POLYMERASE
 (EC 2.7.7.48); HELICASE].
 OS Hepatitis E virus (strain Mexico) (HEV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage.
 OX NCBI_TaxId=31768;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93079857; PubMed=1448913;
 Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fry K.E.,

RA Bradley D.W., Tam A.W., Reyes G.R.;
RT "Molecular cloning and sequencing of the Mexico isolate of hepatitis
RT E virus (HEV)."
RL Virology 191:550-558(1992).
RN (21)
RP SEQUENCE OF 965-1691 FROM N.A.
RX MEDLINE-92271462; PubMed-1589964;
RA Fry K.E., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,
RA Platak M., Feldman R.A., Yun K.Y., Purdy M.A., McCaustland K.A.,
RA Bradley D.W., Reyes G.R.;
RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
RT region encoding consensus motifs for an RNA-dependent RNA polymerase
RT and an ATP/GTP binding site."
RL Virus Genes 6:173-185(1992).
CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
CC ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; M74506; AAA45730.1; -
DR PIR; A44212; A44212.
DR MEROPS; C41.001; -
DR InterPro: IPR000606; -
DR InterPro: IPR002588; -
DR InterPro: IPR002589; -
DR Pfam; PF01661; DUF27; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01660; Vmethyltransf; 1.
KM Polyprotein; Transferase; RNA-directed RNA polymerase; Helicase;
ATP-binding; 973 980 ATP (POTENTIAL).
FT NP_BIND 1691 AA: 185224 MW: DB5F0B2C913F871B CRC64:
SQ SEQUENCE

Query Match 82.5%; Score 7433; DB 1; Length 1691;
Best local similarity 81.0%; Pred. No. 0;
Matches 1384; Conservative 121; Mismatches 166; Indels 38; Gaps 5;

QY 1 PGITTAEQAAALAAANSAANAAVAVVREPLSRVQTEILINLMOPROLYRPREVLNHPRIOR 60
DB 10 PGITTAEQAAALAAANSAANAAVAVVREPLSRVQTEILINLMOPROLYRPREVLNHPRIOR 69
QY 61 VINHELEQYCARAGRCLEVAHNRSTINDNPNVLRHCLRPVGVGVQVQWYAPTRGPAN 120
DB 70 VINHELEQYCARAGRCLEVAHNRSTINDNPNVLRHCLRPVGVGVQVQWYAPTRGPAN 129
QY 121 CRRSALGLPRADRTYCGDFGSCFAAETVALYSLHDLMPADVAEAMRHGTRLYAA 180
DB 130 CRRSALGLPRADRTYCGDFGSCFAAETVALYSLHDLMPADVAEAMRHGTRLYAA 189
QY 181 LHPREVLPRGYHTTSYLLIHGDRAVVTYEGDTSAGYNHDSILFAWTRTKIYGDH 240
DB 190 LHPREVLPRGYHTTSYLLIHGDRAVVTYEGDTSAGYNHDSILFAWTRTKIYGDH 249
QY 241 PLVTERVATGCHVFLLTAAPESPMPYVPRSTEVYVSTFGPGSPSLFSPACSTK 300
DB 250 PLVTERVATGCHVFLLTAAPESPMPYVPRSTEVYVSTFGPGSPSLFSPACSTK 309
QY 301 STEHAVVHIWDRMLMGATLDDAFCSSRLMTYLRGISYVVTGALVANGMNASDAL 360
DB 310 STEHAVVHIWDRMLMGATLDDAFCSSRLMTYLRGISYVVTGALVANGMNASDAL 369
QY 361 TAYTAAYLTICHOYRLTQAIISGMRRLGVEHAOKFTIRLYSMLFEKSGRDYIPGRLO 420
DB 370 TAYTAAYLTICHOYRLTQAIISGMRRLGVEHAOKFTIRLYSMLFEKSGRDYIPGRLO 429
QY 421 FYACGRWLSAGFLDPRVLVFDSEVPCRCRTFLKYAGKFCCEMRMLGQECTCFLEPAE 480

DB 480 FYACGRWLSAGFLDPRVLVFDSEVPCRCRTFLKYAGKFCCEMRMLGQECTCFLEPAE 489
QY 481 GLVGDHGDNEAGSEVDPAEPALDVSCTYAVHGOLEALRYLANVPODIAARASLT 540
DB 490 GLVGDHGDNEAGSEVDPAEPALDVSCTYAVHGOLEALRYLANVPODIAARASLT 549
QY 541 ATVELVASPDLRECRVYLGKNTFRTTVVGAHLAENGPEQVYLSFDSASQSGASHLT 600
DB 550 ATVELVASPDLRECRVYLGKNTFRTTVVGAHLAENGPEQVYLSFDSASQSGASHLT 609
QY 601 YELPAGIQVRISSNGLDCTTFPPGCAAPSAPGEVAAPFCALRYNFTGRHSLTGLM 660
DB 610 YELPAGIQVRISSNGLDCTTFPPGCAAPSAPGEVAAPFCALRYNFTGRHSLTGLM 669
QY 661 LHPGLIGTFPPSGHGWESANPFCGGTLYTRTWSL-----SGFSSDESPPE 709
DB 670 LHPGLIGTFPPSGHGWESANPFCGGTLYTRTWSL-----SGFSSDESPPE 728
QY 710 AAAPAMAATPOLPHSTPVSQDIWVLPPESEFQVDAAPVPAPAPAGLPGVVLTTPPPP 769
DB 729 GGPATATGPAVGSDDSDPD-PLPDVTDGSRPSGAR-PAGPMPNCP----- 774
QY 770 PVHRPSTLPPESRNRLYTYPDGAKVYAGSLPESDCMLVNASNPGRHPPGGGLCHAFYQR 829
DB 775 -----QRRLLHTYPDGAKIYVGSIFESCECTMLVNASNAGHRPPGGGLCHAFYQR 822
QY 830 FPEAFYPTFEFIRREGIAAYLTLPRLIHAVAPDVRVQNPRLAAYRENCSSRRGTAYP 889
DB 823 YPDSFDTAKFVNRDGLAYTLTPRLIHAVAPDVRVQNPRLAAYRENCSSRRGTAYP 882
QY 890 LLGSGIYQVPSLSFDMERNHRBDELYLXLEPAANFEAKPAQVLYLTEDTARTANL 949
DB 883 LLGAGIYQVPSLSFDMERNHRBDELYLXLEPAANFEAKPAQVLYLTEDTARTANL 942
QY 950 ALEIDATEVGRACAGCTISGCIYHIOFTAGVPSSGKSRSIQGGDVVYVPTTELNSM 1009
DB 943 ALEIDSGSEVGRACAGCVERGVYQFTAGVPSSGKSRSIQGGDVVYVPTTELNSM 1002
QY 1010 RRRGFAFTPTTAARVITGRVNVIDEAPSLPRLILLHMORASSVHLLGDPNPAIDFE 1069
DB 1003 RRRGFAFTPTTAARVITGRVNVIDEAPSLPRLILLHMORASSVHLLGDPNPAIDFE 1062
QY 1070 HAGLVPAIRPELAPTSMWXYVTHRCPADVCELIRGAPKRIQTSRYVLSLFWNEBAGOKL 1129
DB 1063 HTGLIPAIRPELAPTSMWXYVTHRCPADVCELIRGAPKRIQTSRYVLSLFWNEBAGOKL 1122
QY 1130 VXTQAAKAAHPGATVHEAGCATETTTTATADARGLIQSSRAHAVALTRHREKCVIL 1189
DB 1123 VXTQAAKAAHPGATVHEAGCATETTTTATADARGLIQSSRAHAVALTRHREKCVIL 1182
QY 1190 DAPGLREVIGSIDVIVNFFLAGGEVGHXRPSTVPRGNPDQNLCTLOAFPSCQISAYHQ 1249
DB 1183 DSPGLREVIGSIDVIVNFFLAGGEVGHXRPSTVPRGNPDQNLCTLOAFPSCQISAYHQ 1242
QY 1250 LAEELGHRPAPVAAVLPCEPELBOGLLYMPOELTVSDSVLEFLTDIVCHMAAPSORKA 1309
DB 1243 LAEELGHRPAPVAAVLPCEPELBOGLLYMPOELTVSDSVLEFLTDIVCHMAAPSORKA 1302
QY 1310 VLSTLVGRYGRRTLYEAHSDVRESLARFTPTIGPQVATTCCELYELVEAVNEAGGOGSA 1369
DB 1303 VLSTLVGRYGRRTLYEAHSDVRESLARFTPTIGPQVATTCCELYELVEAVNEAGGOGSA 1362
QY 1370 VLELDLCNRDVSRTTFQCKXNKFTTETIANGKVAGGQISAMSKTFCALRGPMWRATEKE 1429
DB 1363 VLELDLCNRDVSRTTFQCKXNKFTTETIANGKVAGGQISAMSKTFCALRGPMWRATEKE 1422
QY 1430 ILALLPPIVFGDAYEESVFAAASGAGSCWPFENDFSEPDSTONNPSLGLCYVMEBCG 1489
DB 1423 ILALLPPIVFGDAYEESVFAAASGAGSCWPFENDFSEPDSTONNPSLGLCYVMEBCG 1482
QY 1490 MPQWLIRLHYVRSAMTLOAPKESLKGFWKXHSSEPTILMNTYWNMAIITAHCEPFDNR 1549
DB 1482 MPQWLIRLHYVRSAMTLOAPKESLKGFWKXHSSEPTILMNTYWNMAIITAHCEPFDNR 1549

Db 1483 MPOWLYRLYHAVRSAMIILOAPKESLRFMKHSGSEPSLLMNTVMNMAIIAHCFERDQ 1542
QY 1550 VAAFGDDSVVLCSDYRSGRNAALIIAGCGLLKDYRITGLIAGVYVAPGICITLPDVR 1609
Db 1543 VAAFGDDSVVLCSEYRSGPAGSLIAGGKLLKADFRIDGLIAGVYVAPGALPDVR 1602
QY 1610 FAGRLSEKMGWGPFAEDQLRLAVCDLGLFNVAOVCDVYVRYGVSPGLVHNLIGML 1669
Db 1603 FAGRLSEKMGWGPFAEDQLRLAVCDLGLFNVAOVCDVYVRYGVSPGLVHNLIGML 1662
QY 1670 QTADGKAHFTETIKPVDLTNSIQRYE 1698
Db 1663 QTIGDKAHFTESVAKPILDLTHSIMHRSE 1691
RESULT 4
POLN_HEVY STANDARD: PR1: 1693 AA.
AC 004610:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NON-STRUCTURAL POLYPEPTIDE [CONTAINS: RNA-DIRECTED RNA POLYMERASE
(EC 2.7.7.48); HELICASE].
OS Hepatitis E virus (strain Myanmar) (HEV).
OC Viruses: ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=31769;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93227573; PubMed=8470371;
RA Aye T.T., Uchida T., Ma M.Z., Iida F., Shikata T., Ichikawa M.,
RA Rikhisia T., Wain K.;
RT "Sequence and gene structure of the hepatitis E virus isolated from
Myanmar";
RL Virus genes 7:95-109(1993).
CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: D10330; BAA01172.1; -
DR MEROPS: CA1.001; -
DR InterPro: IPR000606; -
DR InterPro: IPR002588; -
DR InterPro: IPR002589; -
DR Pfam: PF01661; DUF27.1;
DR Pfam: PF01443; Viral_helicase1.1;
DR Pfam: PF01660; Vmethyltransf.1;
DR Polyprotein: Transferase; RNA-directed RNA polymerase; Helicase;
KM ATP-binding.
ET NP_BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185215 MW; AABAC9140AE21EA CRC64;
Query Match 82.5%; Score 7433; DB 1: Length 1693;
Best Local Similarity 81.1%; Pred. No. 0; Mismatches 159; Indels 56; Gaps 6;
Matches 1394; Conservative 109; Mismatches 159; Indels 56; Gaps 6;
QY 2 GITTAIEQAALAAANSALANAVVRFSLRVQTELLINMQPRLVFRPEVLNMFIOQV 61
Db 11 GITTAIEQAALAAANSALANAVVRFSLRVQTELLINMQPRLVFRPEVLNMFIOQV 70
QY 62 IHNELEQYCRAGRCLEAGHPRSINDPNVLRHCFLEPVRDVOYKWTSAFTRGPAAC 121
Db 71 IHNELEQYCRAGRCLEAGHPRSINDPNVLRHCFLEPVRDVOYKWTSAFTRGPAAC 130
QY 122 RRSALRGLEPPARITYCPOFSRCAPAFETGVALYSLHDLMPADVAEAMRHGXTRLYAL 181

Db 131 RRSALRGLEPPARITYCPOFSRCAPAFETGVALYSLHDLMPADVAEAMRHGXTRLYAL 190
QY 182 HLPPEVLLPQGTYYHTSYLLIHDGDAVYVEGDTISAGYNHDVSIIRAWIRTKIYGDHP 241
Db 191 HLPPEVLLPQGTYYHTSYLLIHDGDAVYVEGDTISAGYNHDVSIIRAWIRTKIYGDHP 250
QY 242 LVIERVRAIGCFVLLTLTAPEPSPMPVYPRPRSTEVYRSIFPGGSGSLPSSACTRS 301
Db 251 LVIERVRAIGCFVLLTLTAPEPSPMPVYPRPRSTEVYRSIFPGGSGSLPSSACTRS 310
QY 302 TFAVAVHIMWRLMLFGATLDQAFCCSRMLMTYLRGISKVVAVGALVANEGNASDALT 361
Db 311 TFAVAVHIMWRLMLFGATLDQAFCCSRMLMTYLRGISKVVAVGALVANEGNASDALT 370
QY 362 AXITTAAYLTICQRYLRTQAIKGMRLGVEHAQKFTITRYLSMLPEKSGRDYIPGROLQ 421
Db 371 AXITTAAYLTICQRYLRTQAIKGMRLGVEHAQKFTITRYLSMLPEKSGRDYIPGROLQ 430
QY 422 YACRRMLTSAGHLDPRVLYFDESVPCRCRTFLKVAAGFCFPMRLGDECTCFLEPABG 481
Db 431 YACRRMLTSAGHLDPRVLYFDESVPCRCRTFLKVAAGFCFPMRLGDECTCFLEPABG 490
QY 482 LVGDGHDEAGYEGSEVDPAEPALHDVSGTYAVHGHQLEALYRALNVPODIAARASRLTA 541
Db 491 LVGDGHDEAGYEGSEVDPAEPALHDVSGTYAVHGHQLEALYRALNVPODIAARASRLTA 550
QY 542 TVELVASPDRLCEKRYLVGNKTRTFTVVDGAHLHANGPEQYVLSFPAASQMGAGSHSLTY 601
Db 551 TVKVSQVDRICETLLGNKTRFTSFDVDAVLEANGPERYNIISFASOSTMAAGFESLYT 610
QY 602 ELTPAGLOVRISNGLDCAFTPPGAPSAAPGEVAFCASLYRNRFRORISLQGLML 661
Db 611 AASAGLEVRYAAGLDHRAVFAVSPRSPASAPGEVAFCASLYRNRFRORISLQGLML 670
QY 662 HPEGGLGTPPSPGHIWESANPFCGEGTLYRTWS-TGSEFSDSP-----PEAAAPAM 715
Db 671 HPEGGLGTPPSPGHIWESANPFCGEGTLYRTWS-TGSEFSDSP-----PEAAAPAM 730
QY 716 AATPGLPHSTPVSDIWLVPPESEFQVDAVPAEPDAGLPGVYLTPPPPPYHKPS 775
Db 731 AATPGLPHSTPVSDIWLVPPESEFQVDAVPAEPDAGLPGVYLTPPPPPYHKPS 785
QY 776 TPPP-----SRNRLLYTPPDGAKYVAGSLFESDCDMLVNASNPGRRPG 820
Db 785 TPPP-----SRNRLLYTPPDGAKYVAGSLFESDCDMLVNASNPGRRPG 835
QY 821 GLCHAFYQRFPAFYTFEFTIMREGLAAYTLTPRPIIHAVAPDYRVBONPKRLEAAYREC 880
Db 835 GLCHAFYQRFPAFYTFEFTIMREGLAAYTLTPRPIIHAVAPDYRVBONPKRLEAAYREC 875
QY 881 SRRGTAAVPLLGSIYQVPSLFDAMERNHRRGDELYLTPBAAMWEANKPAQVLYTT 940
Db 896 SRRGTAAVPLLGSIYQVPSLFDAMERNHRRGDELYLTPBAAMWEANKPAQVLYTT 935
QY 941 EDTARTANALALIDATVEGRACAGTISPGIYHVOFTAGVGGSGSKSIQGDVDVYV 1000
Db 956 EDTARTANALALIDATVEGRACAGTISPGIYHVOFTAGVGGSGSKSIQGDVDVYV 995
QY 1001 PIRELNSWRGGAFAFTHTAARVYIGRRVYIDEAPSLPHLLHMOARASSVHLGDP 1060
Db 996 PIRELNSWRGGAFAFTHTAARVYIGRRVYIDEAPSLPHLLHMOARASSVHLGDP 1055
QY 1061 NOIPALIDEHAQVPAIRBELAPTSWMVYTHRCPADVCELIGAVPKIOTTSRVLSLFW 1120
Db 1056 NOIPALIDEHAQVPAIRBELAPTSWMVYTHRCPADVCELIGAVPKIOTTSRVLSLFW 1115
QY 1121 NEPAIGQKLYTHQAAKANPAGIYTHAGAGATFTTTATADARGLIOSSRAHAIYALT 1180
Db 1116 NEPAIGQKLYTHQAAKANPAGIYTHAGAGATFTTTATADARGLIOSSRAHAIYALT 1175
QY 1181 RHTEKCVILADAGLLREVGISDVIVNPFELAGGVGXHRRPSVIRGNPNQNGTQAPRP 1240

Db 1176 RHTEKVEYIIDAPGLIREVIGISDAIVNNFLAGSIGHQREPSVIRPGNDPNDVTLAAPP 1235
 QY 1241 SCOTSAVHOLAEELGHRPAVAVLPPCPPELEGLLYMPELVSDSVLAEFLDIYHCR 1300
 Db 1236 SCOTSAVHOLAEELGHRPAVAVLPPCPPELEGLLYMPELVSDSVLAEFLDIYHCR 1295
 QY 1301 MAAPSORKAVLSTLVGRYGRRTKLYEAHSDVRESLARFPTIGPVATTCETELVEAM 1360
 Db 1296 MAAPNOKAVLSTLVGRYGRRTKLYEAHSDVRESLARFPTIGPVATTCETELVEAM 1355
 QY 1361 VERGODSVALELDLCNRDVSRTTFPOKXCNKFTTGTTIAGKVGQISAMSKTFCLFEG 1420
 Db 1356 VERGODSVALELDLCNRDVSRTTFPOKXCNKFTTGTTIAGKVGQISAMSKTFCLFEG 1415
 QY 1421 PMFRATEKITALLPPIFYGDAAVEESVFAAAGSACWVFENDSEFSTONFSLGL 1480
 Db 1416 PMFRATEKITALLPPIFYGDAAVEESVFAAAGSACWVFENDSEFSTONFSLGL 1475
 QY 1481 ECVYMEECGMPQWILRLYLHVSAMILLQAPRESLKGFWKXSGEPGLLNNVWMAIIA 1540
 Db 1476 ECAIMEECGMPQWILRLYLHVSAMILLQAPRESLKGFWKXSGEPGLLNNVWMAIIA 1535
 QY 1541 HCEFRFRRAAARKGDSVYLCSYRSRMAALLIACCGIKLYDYRPTIGLYAGVYVAPG 1600
 Db 1536 HCEFRFRRAAARKGDSVYLCSYRSRMAALLIACCGIKLYDYRPTIGLYAGVYVAPG 1595
 QY 1601 LGTLPDVFRAGRLSEKNMGPGPERABQLRLAYACDFLGLTNVAQVVDVVSRYGVSPG 1660
 Db 1596 LGTLPDVFRAGRLSEKNMGPGPERABQLRLAYACDFLGLTNVAQVVDVVSRYGVSPG 1655
 QY 1661 LVHNLGMLQTIADGKAHFTETIKPVLDITNSIIQRYE 1698
 Db 1656 LVHNLGMLQTIADGKAHFTETIKPVLDITNSIIQRYE 1693
 RESULT 5
 ID RPO_CGWS STANDARD: PRT: 1648 AA.
 AC P19523; Q83208; P89877; P90356;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-OCT-2000 (Rel. 40, Last sequence update)
 DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (186 KDA PROTEIN) (CONTAINS:
 DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (129 KDA PROTEIN)).
 OS Cucurbit green mottle mosaic virus (watermelon strain SH) (CGMVV), and
 OS Cucurbit green mottle mosaic virus (watermelon strain W) (CGMVV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
 OX NCBI_TaxID=12236; 12237;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN-SH;
 RA Utagaki M., Tomiyama M., Kakutani T., Hidaka S., Kiguchi T., Nagata R.,
 RA Sato T., Motoyoshi F., Nishiguchi M.,
 RT Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP SEQUENCE OF 1496-1648 FROM N.A.
 RC STRAIN-W;
 RX MEDLINE=89073773; PubMed=3201760;
 RA Saito T., Imai Y., Meshi T., Okada Y.,
 RT "Interviral homologies of the 30K proteins of tobamoviruses."
 RT Virology 167:653-656(1988).
 CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
 CC RNA REPLICATION.
 CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA

CC CAPING AND AN RNA HELICASE.
 CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR UAG OCCURS BETWEEN
 CC CODONS FOR LYS-1144 AND GLN-1146.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch.)
 CC
 CC EMBL: D12505; BAA18895.1;
 CC EMBL: D12505; BAA18896.1;
 CC EMBL: J04322; AAA46382.1;
 CC PIR: J01157; WMTMS2;
 CC InterPro: IPR000606;
 CC InterPro: IPR001788;
 CC InterPro: IPR002588;
 CC Pfam: PF00978; RNA_dep_RNAPol2; 1.
 CC Pfam: PF01443; Viral_helicase1; 1.
 CC Pfam: PF01660; Ymethyltransf; 1.
 CC
 CC Transferase: RNA-directed RNA polymerase; Helicase; ATP-binding.
 CC FT CHAIN 1 1648 RNA-DIRECTED RNA POLYMERASE.
 CC FT NP_BIND 1 1144 METHYLTRANSFERASE/RNA HELICASE.
 CC FT NP_BIND 863 870 ATP (POTENTIAL).
 CC SEQUENCE 1648 AA; 186549 MW; 1D1AFEEB7B65595 CRC64;
 Query Match 4.5%; Score 406.5; DB 1; Length 1648;
 Best Local Similarity 18.8%; Pred. No. 6,7e-18;
 Matches 350; Conservative 220; Mismatches 633; Indels 661; Gaps 73;
 QY 77 CLEVGA-HPRSDNPNVLRRC---FLRPVGRDYOQWYSAPRGPANCRSALGLP- 130
 Db 105 CYDIGNYTOHLKRGSYVHCNCPCLDKVARNVM-YNDMTTOHYORHSGSGCRLPT 163
 QY 131 ---PADRTY-----CFDGFSCAFMAETG---VALYSLHDIIMPADVAMARHG 174
 Db 164 FQIDAFRRYDSSPCAVTSDVDFGCSYDFSGRDNHNAVLSHSYDIPIYSIGALHRKNV 223
 QY 175 TRIYALHLRPEVLL--PRTYITTSYLLIHDDRAVYVTEGTSAGYNIDVSLRWIR 232
 Db 224 RVCYAAHFSEALLLSSPVGNINLSIGAQFRVDDDVHFLFSESTLHYHSLNKLIVM 283
 QY 223 TRYIVDHPVIER---VRAIGCFVLLTLAEPSPMPVPPYPRSTEVYVRSIFPGGS 289
 Db 284 RTYFPADDRVYIKEMWKRVDFPFRFLVADTHMLHKSIGHYSKSKSEFAL-----NT 338
 QY 290 PSLFSPACSTKSTF-----HAPVHW 311
 Db 339 PRIF-----QDKATFSVWPEAKRKVLIPKPELSRFLSGNVKISRMLVDADFVHTIINH- 393
 QY 312 DRMLLEGATIDDOAFCCSRMTYLRGISYVYVYGVYALVANGMNASDALTAITAAVLT 371
 Db 394 -----STYDNKALVWKVKNQVSEVRSIRSVYVGVSEWNVVPVQDLDISFIFLLV 446
 QY 372 CHQRYLRTQAIKSGMRGLGVEHQAQKFTIRLYSWLFEKSGNDYIPGRLODYAACRMLSA 431
 Db 447 -----KVRKVOIE----- 454
 QY 432 GFHLDRVLVFEDESVPCRTFLKLVAGKFCFCFMRMLGQECCTFLPBAELVGDHGDNE 491
 Db 455 -----LMSDKVYIEARGLLRFRFADSLKSAVEGLG-DCV-----YDALVGTGWFDS 499
 QY 492 AYGESEYDAPEPAHLVDVSGYVAHGHOLEALYRALVNPQDIARASRLATVYLVASPD 551
 Db 500 SDELKVLPL-EP-----FMFESDYLDEMYEA-----DAKIRRESVS-ELIASGDD 542
 QY 552 L-----ECRVLVANKTFR-----TVVQGHLEANGPQYVLSFPASQSMG 593
 Db 543 LFKKIDIEIRNNSGVFEVDKPEQFCKELNVNPMILIGHVTEA-----IFSOKAGVTVTG 596

```

OY 594 AGSHSLTELEFP-AGLOYRISNSGLD-C-----TATPPGAGDAPSAARCEVAARCS 641
DB 597 LFT-----LSPEMGASVALSTSVDTCEMDVTEMDIYLMADKSHSYMSPEMARWMD 650
OY 642 ALYRYNRFORHSLTGLMLHPEGLGIFPPFSPGHIMESANPFCGEGTLRYTSTSGF 701
DB 651 VYKGNK-----GALVEYKVGTSMTLP-----ATAWENG- 679
OY 702 SDDSPPEAPAMAAATPGLPHSTPPVDIWLPP-----PSEE-----FOVDA 746
DB 680 -----KAVLPGLSICVRRKPOFSKPLDEEDDLRLSNMNFKVS 718
OY 747 PVPAPADAGLPVLPPLPPPPVHKPSIIPPSNRRLTYTPDGAKYAGS----- 799
DB 719 KUKKITTTPVYVTGTI-----REROMKMYIDYLSASLGSTGJNER 758
OY 800 LEESD-----CDMLYNASNPGRPGGGLCHAFYQREPEAFYPTFTM 841
DB 759 IYRSDMNGTEESMOTFGLYDCEKCKMLLPKAKHMAVYLA-----SDDTT 805
OY 842 REGLAAYTLTPPTIHAAPDYRVQONPKRELEAATRETCRRGTAAVPLSGITQVPS 901
DB 806 RIFFLSYDESGSPID-----KKNMRF-----TKVYSVIRS 842
OY 902 LSFDAWERNHRRGDELYLEPAANWFEANKPAQPVLTITEDTARTANLALIDAATEVGR 961
DB 843 LEV-----LNKEA----- 850
OY 963 ACAGCTISPIVHYQTAGVPGSGKRSI---OQGDVYVVPVTRLRNSMRRGFAFT 1018
DB 851 -----IVDPG-VHITLVGDVPCGKTAETIARVNMTDLVLPGRBAAMIRACALIK 904
OY 1019 PHTAAR-----VTIGRR-----VVIDEAPSLPPLL--LHMKORASSVHLGDPNQ 1062
DB 905 SPVANDNDVRTDSEFMNKKIKFPAVYVDEGLMVTGLMLNALKISGCKKAFVVEDAKO 964
OY 1063 IPAT-----DEHAGLVPAIRPELAPTSWMXVTHRCPADVCILRGAY-PKIGTTSVLNS 1117
DB 965 IPIFRVMNFDPKELRFLIVDNVERRY--VTHRCRDVTSPLNTYKAAVATTSVNS 1022
OY 1118 LEWNEPAI-----GOKLVYQOA-----AKAANPGATVHEAQAFTFT 1156
DB 1023 V-----KAIVSGAGILRPETLKIKGIITFTOSDKSLKSGYNDVNYHELOGEFEET 1078
OY 1157 TII-ATADAGLIQSSRAHAIVALTFRHEK-----CVIIDAPGLREVGISDYVNNFELA 1211
DB 1079 AVVRAATPPIGLIARDSPHVALTRHTKAMYYTVFEDAVNSI---IADV----- 1126
OY 1212 GGEVGHKRPVPIPRGNPQONLGTLOAFPSCQISAVHQAELGHRPAVAVALPPCPRL 1271
DB 1127 -----EKVDSILTM-----FAATVPKXOL 1147
OY 1272 EGGLEYMEQLTVSDSVLYFELTDIVHCHMAPSORKAVLSTLVGRYGR----- 1320
DB 1148 MONSLY-----VHRNIFLPVSKGTGYTMOEFYDKCLGNSFVLND 1188
OY 1321 -----RTKLYEAHSDVRESIARFTPT-----IGPVQATCE-----LYELV 1357
DB 1189 FDAVTMRLELDNEFNLOPCRILTSNEDPVALIKNEAQNFLIVLTACRPRIPGLLENT 1248
OY 1358 EAMVEKGODGSAVL--ELDLCNRDVS-----RTTFQKCKNKT 1394
DB 1249 VAMIRNMNTPLDAGTVDTNMSISIVDNFSSSVRDEVLDLHDCVRASSIQSFSWMS 1308
OY 1395 TGETTANGKVG-----OGISAMSKTFCAL 1418
DB 1309 CQPTSAVGOGLANFNFLDPAFTYMHMIKROKSRDLNISIOSEYALQTIYVHPKYNAV 1368
OY 1419 FGPWFRAIEKELTALLPBN--IFYGDAVEESY--FAAAYSAGSGCMVFENDESDSTON 1474
DB 1369 FGPVPEKYTLTKFLSMVSDSKFFPYTRKKRPEDLOEFFSDSHSDYEILELDVSKYDKSOS 1428
OY 1475 NFSJGLECVMEGCMQMLIRFLYLVRSAMWLTQAPKESLKG--FMKKHSGSPGILLMNT 1532

```

```

DB 1429 DPHSEIEMAIWEKLDLDILAMMSMGKRTTLQDPQAGIKRLIYQKSGSVTTFIGNT 1488
OY 1533 -----VNMALIAHCYEFERFVAFAFGDDSVLCS---DYRQSRMAALTAGGLKIK 1583
DB 1489 FIACVASMPLPDKCFK-----ASFGDDSLIYLPGLLEVPDIQAAVAVL--WNFEAK 1540
OY 1584 VDIPIGLYACVYV---APGLCTLPDYVRFAGRLSEKMMGPPEAEQRLAVCFPLKGL 1640
DB 1541 LFRKRYGYFCGKYIIHHANGCIVPDPPLKLSKLGKNSL-VGEYHEVEFRISLIDVAHSL 1599
OY 1641 TNVA 1644
DB 1600 FNGA 1603

RESULT 6
POLN_RUBVT STANDARD: PRF: 2205 AA.
AC P13889;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-AUG-1992 (Rel. 23, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE [CONTAINS: NONSTRUCTURAL PROTEINS NSP1 TO NSP4].
OS Rubella virus (strain Thertien).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Togaviridae;
OC Rubivirus.
OX NCBI_TaxID=11045;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90281585; PubMed=2353453;
RA Dominguez G., Wang C.Y., Frey T.K.;
RT "Sequence of the genome RNA of rubella virus: evidence for genetic rearrangement during togavirus evolution.";
RL Virology 177:225-258(1990).
RN [2]
RP SEQUENCE OF 1737-2205 FROM N.A.
RX MEDLINE=88226020; PubMed=2836271;
RA Frey T.K., Marz L.D.;
RT "Sequence of the region coding for virion proteins C and E2 and the carboxy terminus of the nonstructural proteins of rubella virus: comparison with alphaviruses.";
RL Gene 62:85-99(1988).
CC -I- P1M: SPECIFIC ENZYMOLOGICAL CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb.ch/announce/ or send an email to license@isb-sib.ch).
CC -----
DB EMBL: M15240; AA88528.1; -
DR PIR: A35320; MNWVN.
DR MEROPS: C27.001; -
DR InterPro: IPR002589; -
DR Pfam: PF01661; DUF27.1.
KW Polypeptide; Nonstructural protein.
FT CHAIN 1 ? NONSTRUCTURAL PROTEIN NSP1.
FT CHAIN 2 ? NONSTRUCTURAL PROTEIN NSP2.
FT CHAIN 3 ? NONSTRUCTURAL PROTEIN NSP3.
FT CHAIN 4 ? NONSTRUCTURAL PROTEIN NSP4.
FT CHAIN 5 ? NONSTRUCTURAL PROTEIN NSP5.
FT CHAIN 6 ? NONSTRUCTURAL PROTEIN NSP6.
FT CHAIN 7 ? NONSTRUCTURAL PROTEIN NSP7.
FT CHAIN 8 ? NONSTRUCTURAL PROTEIN NSP8.
FT CHAIN 9 ? NONSTRUCTURAL PROTEIN NSP9.
FT CHAIN 10 ? NONSTRUCTURAL PROTEIN NSP10.
FT CHAIN 11 ? NONSTRUCTURAL PROTEIN NSP11.
FT CHAIN 12 ? NONSTRUCTURAL PROTEIN NSP12.
FT CHAIN 13 ? NONSTRUCTURAL PROTEIN NSP13.
FT CHAIN 14 ? NONSTRUCTURAL PROTEIN NSP14.
FT CHAIN 15 ? NONSTRUCTURAL PROTEIN NSP15.
FT CHAIN 16 ? NONSTRUCTURAL PROTEIN NSP16.
FT CHAIN 17 ? NONSTRUCTURAL PROTEIN NSP17.
FT CHAIN 18 ? NONSTRUCTURAL PROTEIN NSP18.
FT CHAIN 19 ? NONSTRUCTURAL PROTEIN NSP19.
FT CHAIN 20 ? NONSTRUCTURAL PROTEIN NSP20.
FT CHAIN 21 ? NONSTRUCTURAL PROTEIN NSP21.
FT CHAIN 22 ? NONSTRUCTURAL PROTEIN NSP22.
FT CHAIN 23 ? NONSTRUCTURAL PROTEIN NSP23.
FT CHAIN 24 ? NONSTRUCTURAL PROTEIN NSP24.
FT CHAIN 25 ? NONSTRUCTURAL PROTEIN NSP25.
FT CHAIN 26 ? NONSTRUCTURAL PROTEIN NSP26.
FT CHAIN 27 ? NONSTRUCTURAL PROTEIN NSP27.
FT CHAIN 28 ? NONSTRUCTURAL PROTEIN NSP28.
FT CHAIN 29 ? NONSTRUCTURAL PROTEIN NSP29.
FT CHAIN 30 ? NONSTRUCTURAL PROTEIN NSP30.
FT CHAIN 31 ? NONSTRUCTURAL PROTEIN NSP31.
FT CHAIN 32 ? NONSTRUCTURAL PROTEIN NSP32.
FT CHAIN 33 ? NONSTRUCTURAL PROTEIN NSP33.
FT CHAIN 34 ? NONSTRUCTURAL PROTEIN NSP34.
FT CHAIN 35 ? NONSTRUCTURAL PROTEIN NSP35.
FT CHAIN 36 ? NONSTRUCTURAL PROTEIN NSP36.
FT CHAIN 37 ? NONSTRUCTURAL PROTEIN NSP37.
FT CHAIN 38 ? NONSTRUCTURAL PROTEIN NSP38.
FT CHAIN 39 ? NONSTRUCTURAL PROTEIN NSP39.
FT CHAIN 40 ? NONSTRUCTURAL PROTEIN NSP40.
FT CHAIN 41 ? NONSTRUCTURAL PROTEIN NSP41.
FT CHAIN 42 ? NONSTRUCTURAL PROTEIN NSP42.
FT CHAIN 43 ? NONSTRUCTURAL PROTEIN NSP43.
FT CHAIN 44 ? NONSTRUCTURAL PROTEIN NSP44.
FT CHAIN 45 ? NONSTRUCTURAL PROTEIN NSP45.
FT CHAIN 46 ? NONSTRUCTURAL PROTEIN NSP46.
FT CHAIN 47 ? NONSTRUCTURAL PROTEIN NSP47.
FT CHAIN 48 ? NONSTRUCTURAL PROTEIN NSP48.
FT CHAIN 49 ? NONSTRUCTURAL PROTEIN NSP49.
FT CHAIN 50 ? NONSTRUCTURAL PROTEIN NSP50.
FT CHAIN 51 ? NONSTRUCTURAL PROTEIN NSP51.
FT CHAIN 52 ? NONSTRUCTURAL PROTEIN NSP52.
FT CHAIN 53 ? NONSTRUCTURAL PROTEIN NSP53.
FT CHAIN 54 ? NONSTRUCTURAL PROTEIN NSP54.
FT CHAIN 55 ? NONSTRUCTURAL PROTEIN NSP55.
FT CHAIN 56 ? NONSTRUCTURAL PROTEIN NSP56.
FT CHAIN 57 ? NONSTRUCTURAL PROTEIN NSP57.
FT CHAIN 58 ? NONSTRUCTURAL PROTEIN NSP58.
FT CHAIN 59 ? NONSTRUCTURAL PROTEIN NSP59.
FT CHAIN 60 ? NONSTRUCTURAL PROTEIN NSP60.
FT CHAIN 61 ? NONSTRUCTURAL PROTEIN NSP61.
FT CHAIN 62 ? NONSTRUCTURAL PROTEIN NSP62.
FT CHAIN 63 ? NONSTRUCTURAL PROTEIN NSP63.
FT CHAIN 64 ? NONSTRUCTURAL PROTEIN NSP64.
FT CHAIN 65 ? NONSTRUCTURAL PROTEIN NSP65.
FT CHAIN 66 ? NONSTRUCTURAL PROTEIN NSP66.
FT CHAIN 67 ? NONSTRUCTURAL PROTEIN NSP67.
FT CHAIN 68 ? NONSTRUCTURAL PROTEIN NSP68.
FT CHAIN 69 ? NONSTRUCTURAL PROTEIN NSP69.
FT CHAIN 70 ? NONSTRUCTURAL PROTEIN NSP70.
FT CHAIN 71 ? NONSTRUCTURAL PROTEIN NSP71.
FT CHAIN 72 ? NONSTRUCTURAL PROTEIN NSP72.
FT CHAIN 73 ? NONSTRUCTURAL PROTEIN NSP73.
FT CHAIN 74 ? NONSTRUCTURAL PROTEIN NSP74.
FT CHAIN 75 ? NONSTRUCTURAL PROTEIN NSP75.
FT CHAIN 76 ? NONSTRUCTURAL PROTEIN NSP76.
FT CHAIN 77 ? NONSTRUCTURAL PROTEIN NSP77.
FT CHAIN 78 ? NONSTRUCTURAL PROTEIN NSP78.
FT CHAIN 79 ? NONSTRUCTURAL PROTEIN NSP79.
FT CHAIN 80 ? NONSTRUCTURAL PROTEIN NSP80.
FT CHAIN 81 ? NONSTRUCTURAL PROTEIN NSP81.
FT CHAIN 82 ? NONSTRUCTURAL PROTEIN NSP82.
FT CHAIN 83 ? NONSTRUCTURAL PROTEIN NSP83.
FT CHAIN 84 ? NONSTRUCTURAL PROTEIN NSP84.
FT CHAIN 85 ? NONSTRUCTURAL PROTEIN NSP85.
FT CHAIN 86 ? NONSTRUCTURAL PROTEIN NSP86.
FT CHAIN 87 ? NONSTRUCTURAL PROTEIN NSP87.
FT CHAIN 88 ? NONSTRUCTURAL PROTEIN NSP88.
FT CHAIN 89 ? NONSTRUCTURAL PROTEIN NSP89.
FT CHAIN 90 ? NONSTRUCTURAL PROTEIN NSP90.
FT CHAIN 91 ? NONSTRUCTURAL PROTEIN NSP91.
FT CHAIN 92 ? NONSTRUCTURAL PROTEIN NSP92.
FT CHAIN 93 ? NONSTRUCTURAL PROTEIN NSP93.
FT CHAIN 94 ? NONSTRUCTURAL PROTEIN NSP94.
FT CHAIN 95 ? NONSTRUCTURAL PROTEIN NSP95.
FT CHAIN 96 ? NONSTRUCTURAL PROTEIN NSP96.
FT CHAIN 97 ? NONSTRUCTURAL PROTEIN NSP97.
FT CHAIN 98 ? NONSTRUCTURAL PROTEIN NSP98.
FT CHAIN 99 ? NONSTRUCTURAL PROTEIN NSP99.
FT CHAIN 100 ? NONSTRUCTURAL PROTEIN NSP100.
FT CHAIN 101 ? NONSTRUCTURAL PROTEIN NSP101.
FT CHAIN 102 ? NONSTRUCTURAL PROTEIN NSP102.
FT CHAIN 103 ? NONSTRUCTURAL PROTEIN NSP103.
FT CHAIN 104 ? NONSTRUCTURAL PROTEIN NSP104.
FT CHAIN 105 ? NONSTRUCTURAL PROTEIN NSP105.
FT CHAIN 106 ? NONSTRUCTURAL PROTEIN NSP106.
FT CHAIN 107 ? NONSTRUCTURAL PROTEIN NSP107.
FT CHAIN 108 ? NONSTRUCTURAL PROTEIN NSP108.
FT CHAIN 109 ? NONSTRUCTURAL PROTEIN NSP109.
FT CHAIN 110 ? NONSTRUCTURAL PROTEIN NSP110.
FT CHAIN 111 ? NONSTRUCTURAL PROTEIN NSP111.
FT CHAIN 112 ? NONSTRUCTURAL PROTEIN NSP112.
FT CHAIN 113 ? NONSTRUCTURAL PROTEIN NSP113.
FT CHAIN 114 ? NONSTRUCTURAL PROTEIN NSP114.
FT CHAIN 115 ? NONSTRUCTURAL PROTEIN NSP115.
FT CHAIN 116 ? NONSTRUCTURAL PROTEIN NSP116.
FT CHAIN 117 ? NONSTRUCTURAL PROTEIN NSP117.
FT CHAIN 118 ? NONSTRUCTURAL PROTEIN NSP118.
FT CHAIN 119 ? NONSTRUCTURAL PROTEIN NSP119.
FT CHAIN 120 ? NONSTRUCTURAL PROTEIN NSP120.
FT CHAIN 121 ? NONSTRUCTURAL PROTEIN NSP121.
FT CHAIN 122 ? NONSTRUCTURAL PROTEIN NSP122.
FT CHAIN 123 ? NONSTRUCTURAL PROTEIN NSP123.
FT CHAIN 124 ? NONSTRUCTURAL PROTEIN NSP124.
FT CHAIN 125 ? NONSTRUCTURAL PROTEIN NSP125.
FT CHAIN 126 ? NONSTRUCTURAL PROTEIN NSP126.
FT CHAIN 127 ? NONSTRUCTURAL PROTEIN NSP127.
FT CHAIN 128 ? NONSTRUCTURAL PROTEIN NSP128.
FT CHAIN 129 ? NONSTRUCTURAL PROTEIN NSP129.
FT CHAIN 130 ? NONSTRUCTURAL PROTEIN NSP130.
FT CHAIN 131 ? NONSTRUCTURAL PROTEIN NSP131.
FT CHAIN 132 ? NONSTRUCTURAL PROTEIN NSP132.
FT CHAIN 133 ? NONSTRUCTURAL PROTEIN NSP133.
FT CHAIN 134 ? NONSTRUCTURAL PROTEIN NSP134.
FT CHAIN 135 ? NONSTRUCTURAL PROTEIN NSP135.
FT CHAIN 136 ? NONSTRUCTURAL PROTEIN NSP136.
FT CHAIN 137 ? NONSTRUCTURAL PROTEIN NSP137.
FT CHAIN 138 ? NONSTRUCTURAL PROTEIN NSP138.
FT CHAIN 139 ? NONSTRUCTURAL PROTEIN NSP139.
FT CHAIN 140 ? NONSTRUCTURAL PROTEIN NSP140.
FT CHAIN 141 ? NONSTRUCTURAL PROTEIN NSP141.
FT CHAIN 142 ? NONSTRUCTURAL PROTEIN NSP142.
FT CHAIN 143 ? NONSTRUCTURAL PROTEIN NSP143.
FT CHAIN 144 ? NONSTRUCTURAL PROTEIN NSP144.
FT CHAIN 145 ? NONSTRUCTURAL PROTEIN NSP145.
FT CHAIN 146 ? NONSTRUCTURAL PROTEIN NSP146.
FT CHAIN 147 ? NONSTRUCTURAL PROTEIN NSP147.
FT CHAIN 148 ? NONSTRUCTURAL PROTEIN NSP148.
FT CHAIN 149 ? NONSTRUCTURAL PROTEIN NSP149.
FT CHAIN 150 ? NONSTRUCTURAL PROTEIN NSP150.
FT CHAIN 151 ? NONSTRUCTURAL PROTEIN NSP151.
FT CHAIN 152 ? NONSTRUCTURAL PROTEIN NSP152.
FT CHAIN 153 ? NONSTRUCTURAL PROTEIN NSP153.
FT CHAIN 154 ? NONSTRUCTURAL PROTEIN NSP154.
FT CHAIN 155 ? NONSTRUCTURAL PROTEIN NSP155.
FT CHAIN 156 ? NONSTRUCTURAL PROTEIN NSP156.
FT CHAIN 157 ? NONSTRUCTURAL PROTEIN NSP157.
FT CHAIN 158 ? NONSTRUCTURAL PROTEIN NSP158.
FT CHAIN 159 ? NONSTRUCTURAL PROTEIN NSP159.
FT CHAIN 160 ? NONSTRUCTURAL PROTEIN NSP160.
FT CHAIN 161 ? NONSTRUCTURAL PROTEIN NSP161.
FT CHAIN 162 ? NONSTRUCTURAL PROTEIN NSP162.
FT CHAIN 163 ? NONSTRUCTURAL PROTEIN NSP163.
FT CHAIN 164 ? NONSTRUCTURAL PROTEIN NSP164.
FT CHAIN 165 ? NONSTRUCTURAL PROTEIN NSP165.
FT CHAIN 166 ? NONSTRUCTURAL PROTEIN NSP166.
FT CHAIN 167 ? NONSTRUCTURAL PROTEIN NSP167.
FT CHAIN 168 ? NONSTRUCTURAL PROTEIN NSP168.
FT CHAIN 169 ? NONSTRUCTURAL PROTEIN NSP169.
FT CHAIN 170 ? NONSTRUCTURAL PROTEIN NSP170.
FT CHAIN 171 ? NONSTRUCTURAL PROTEIN NSP171.
FT CHAIN 172 ? NONSTRUCTURAL PROTEIN NSP172.
FT CHAIN 173 ? NONSTRUCTURAL PROTEIN NSP173.
FT CHAIN 174 ? NONSTRUCTURAL PROTEIN NSP174.
FT CHAIN 175 ? NONSTRUCTURAL PROTEIN NSP175.
FT CHAIN 176 ? NONSTRUCTURAL PROTEIN NSP176.
FT CHAIN 177 ? NONSTRUCTURAL PROTEIN NSP177.
FT CHAIN 178 ? NONSTRUCTURAL PROTEIN NSP178.
FT CHAIN 179 ? NONSTRUCTURAL PROTEIN NSP179.
FT CHAIN 180 ? NONSTRUCTURAL PROTEIN NSP180.
FT CHAIN 181 ? NONSTRUCTURAL PROTEIN NSP181.
FT CHAIN 182 ? NONSTRUCTURAL PROTEIN NSP182.
FT CHAIN 183 ? NONSTRUCTURAL PROTEIN NSP183.
FT CHAIN 184 ? NONSTRUCTURAL PROTEIN NSP184.
FT CHAIN 185 ? NONSTRUCTURAL PROTEIN NSP185.
FT CHAIN 186 ? NONSTRUCTURAL PROTEIN NSP186.
FT CHAIN 187 ? NONSTRUCTURAL PROTEIN NSP187.
FT CHAIN 188 ? NONSTRUCTURAL PROTEIN NSP188.
FT CHAIN 189 ? NONSTRUCTURAL PROTEIN NSP189.
FT CHAIN 190 ? NONSTRUCTURAL PROTEIN NSP190.
FT CHAIN 191 ? NONSTRUCTURAL PROTEIN NSP191.
FT CHAIN 192 ? NONSTRUCTURAL PROTEIN NSP192.
FT CHAIN 193 ? NONSTRUCTURAL PROTEIN NSP193.
FT CHAIN 194 ? NONSTRUCTURAL PROTEIN NSP194.
FT CHAIN 195 ? NONSTRUCTURAL PROTEIN NSP195.
FT CHAIN 196 ? NONSTRUCTURAL PROTEIN NSP196.
FT CHAIN 197 ? NONSTRUCTURAL PROTEIN NSP197.
FT CHAIN 198 ? NONSTRUCTURAL PROTEIN NSP198.
FT CHAIN 199 ? NONSTRUCTURAL PROTEIN NSP199.
FT CHAIN 200 ? NONSTRUCTURAL PROTEIN NSP200.
FT CHAIN 201 ? NONSTRUCTURAL PROTEIN NSP201.
FT CHAIN 202 ? NONSTRUCTURAL PROTEIN NSP202.
FT CHAIN 203 ? NONSTRUCTURAL PROTEIN NSP203.
FT CHAIN 204 ? NONSTRUCTURAL PROTEIN NSP204.
FT CHAIN 205 ? NONSTRUCTURAL PROTEIN NSP205.
FT CHAIN 206 ? NONSTRUCTURAL PROTEIN NSP206.
FT CHAIN 207 ? NONSTRUCTURAL PROTEIN NSP207.
FT CHAIN 208 ? NONSTRUCTURAL PROTEIN NSP208.
FT CHAIN 209 ? NONSTRUCTURAL PROTEIN NSP209.
FT CHAIN 210 ? NONSTRUCTURAL PROTEIN NSP210.
FT CHAIN 211 ? NONSTRUCTURAL PROTEIN NSP211.
FT CHAIN 212 ? NONSTRUCTURAL PROTEIN NSP212.
FT CHAIN 213 ? NONSTRUCTURAL PROTEIN NSP213.
FT CHAIN 214 ? NONSTRUCTURAL PROTEIN NSP214.
FT CHAIN 215 ? NONSTRUCTURAL PROTEIN NSP215.
FT CHAIN 216 ? NONSTRUCTURAL PROTEIN NSP216.
FT CHAIN 217 ? NONSTRUCTURAL PROTEIN NSP217.
FT CHAIN 218 ? NONSTRUCTURAL PROTEIN NSP218.
FT CHAIN 219 ? NONSTRUCTURAL PROTEIN NSP219.
FT CHAIN 220 ? NONSTRUCTURAL PROTEIN NSP220.
FT CHAIN 221 ? NONSTRUCTURAL PROTEIN NSP221.
FT CHAIN 222 ? NONSTRUCTURAL PROTEIN NSP222.
FT CHAIN 223 ? NONSTRUCTURAL PROTEIN NSP223.
FT CHAIN 224 ? NONSTRUCTURAL PROTEIN NSP224.
FT CHAIN 225 ? NONSTRUCTURAL PROTEIN NSP225.
FT CHAIN 226 ? NONSTRUCTURAL PROTEIN NSP226.
FT CHAIN 227 ? NONSTRUCTURAL PROTEIN NSP227.
FT CHAIN 228 ? NONSTRUCTURAL PROTEIN NSP228.
FT CHAIN 229 ? NONSTRUCTURAL PROTEIN NSP229.
FT CHAIN 230 ? NONSTRUCTURAL PROTEIN NSP230.
FT CHAIN 231 ? NONSTRUCTURAL PROTEIN NSP231.
FT CHAIN 232 ? NONSTRUCTURAL PROTEIN NSP232.
FT CHAIN 233 ? NONSTRUCTURAL PROTEIN NSP233.
FT CHAIN 234 ? NONSTRUCTURAL PROTEIN NSP234.
FT CHAIN 235 ? NONSTRUCTURAL PROTEIN NSP235.
FT CHAIN 236 ? NONSTRUCTURAL PROTEIN NSP236.
FT CHAIN 237 ? NONSTRUCTURAL PROTEIN NSP237.
FT CHAIN 238 ? NONSTRUCTURAL PROTEIN NSP238.
FT CHAIN 239 ? NONSTRUCTURAL PROTEIN NSP239.
FT CHAIN 240 ? NONSTRUCTURAL PROTEIN NSP240.
FT CHAIN 241 ? NONSTRUCTURAL PROTEIN NSP241.
FT CHAIN 242 ? NONSTRUCTURAL PROTEIN NSP242.
FT CHAIN 243 ? NONSTRUCTURAL PROTEIN NSP243.
FT CHAIN 244 ? NONSTRUCTURAL PROTEIN NSP244.
FT CHAIN 245 ? NONSTRUCTURAL PROTEIN NSP245.
FT CHAIN 246 ? NONSTRUCTURAL PROTEIN NSP246.
FT CHAIN 247 ? NONSTRUCTURAL PROTEIN NSP247.
FT CHAIN 248 ? NONSTRUCTURAL PROTEIN NSP248.
FT CHAIN 249 ? NONSTRUCTURAL PROTEIN NSP249.
FT CHAIN 250 ? NONSTRUCTURAL PROTEIN NSP250.
FT CHAIN 251 ? NONSTRUCTURAL PROTEIN NSP251.
FT CHAIN 252 ? NONSTRUCTURAL PROTEIN NSP252.
FT CHAIN 253 ? NONSTRUCTURAL PROTEIN NSP253.
FT CHAIN 254 ? NONSTRUCTURAL PROTEIN NSP254.
FT CHAIN 255 ? NONSTRUCTURAL PROTEIN NSP255.
FT CHAIN 256 ? NONSTRUCTURAL PROTEIN NSP256.
FT CHAIN 257 ? NONSTRUCTURAL PROTEIN NSP257.
FT CHAIN 258 ? NONSTRUCTURAL PROTEIN NSP258.
FT CHAIN 259 ? NONSTRUCTURAL PROTEIN NSP259.
FT CHAIN 260 ? NONSTRUCTURAL PROTEIN NSP260.
FT CHAIN 261 ? NONSTRUCTURAL PROTEIN NSP261.
FT CHAIN 262 ? NONSTRUCTURAL PROTEIN NSP262.
FT CHAIN 263 ? NONSTRUCTURAL PROTEIN NSP263.
FT CHAIN 264 ? NONSTRUCTURAL PROTEIN NSP264.
FT CHAIN 265 ? NONSTRUCTURAL PROTEIN NSP265.
FT CHAIN 266 ? NONSTRUCTURAL PROTEIN NSP266.
FT CHAIN 267 ? NONSTRUCTURAL PROTEIN NSP267.
FT CHAIN 268 ? NONSTRUCTURAL PROTEIN NSP268.
FT CHAIN 269 ? NONSTRUCTURAL PROTEIN NSP269.
FT CHAIN 270 ? NONSTRUCTURAL PROTEIN NSP270.
FT CHAIN 271 ? NONSTRUCTURAL PROTEIN NSP271.
FT CHAIN 272 ? NONSTRUCTURAL PROTEIN NSP272.
FT CHAIN 273 ? NONSTRUCTURAL PROTEIN NSP273.
FT CHAIN 274 ? NONSTRUCTURAL PROTEIN NSP274.
FT CHAIN 275 ? NONSTRUCTURAL PROTEIN NSP275.
FT CHAIN 276 ? NONSTRUCTURAL PROTEIN NSP276.
FT CHAIN 277 ? NONSTRUCTURAL PROTEIN NSP277.
FT CHAIN 278 ? NONSTRUCTURAL PROTEIN NSP278.
FT CHAIN 279 ? NONSTRUCTURAL PROTEIN NSP279.
FT CHAIN 280 ? NONSTRUCTURAL PROTEIN NSP280.
FT CHAIN 281 ? NONSTRUCTURAL PROTEIN NSP281.
FT CHAIN 282 ? NONSTRUCTURAL PROTEIN NSP282.
FT CHAIN 283 ? NONSTRUCTURAL PROTEIN NSP283.
FT CHAIN 284 ? NONSTRUCTURAL PROTEIN NSP284.
FT CHAIN 285 ? NONSTRUCTURAL PROTEIN NSP285.
FT CHAIN 286 ? NONSTRUCTURAL PROTEIN NSP286.
FT CHAIN 287 ? NONSTRUCTURAL PROTEIN NSP287.
FT CHAIN 288 ? NONSTRUCTURAL PROTEIN NSP288.
FT CHAIN 289 ? NONSTRUCTURAL PROTEIN NSP289.
FT CHAIN 290 ? NONSTRUCTURAL PROTEIN NSP290.
FT CHAIN 291 ? NONSTRUCTURAL PROTEIN NSP291.
FT CHAIN 292 ? NONSTRUCTURAL PROTEIN NSP292.
FT CHAIN 293 ? NONSTRUCTURAL PROTEIN NSP293.
FT CHAIN 294 ? NONSTRUCTURAL PROTEIN NSP294.
FT CHAIN 295 ? NONSTRUCTURAL PROTEIN NSP295.
FT CHAIN 296 ? NONSTRUCTURAL PROTEIN NSP296.
FT CHAIN 297 ? NONSTRUCTURAL PROTEIN NSP297.
FT CHAIN 298 ? NONSTRUCTURAL PROTEIN NSP298.
FT CHAIN 299 ? NONSTRUCTURAL PROTEIN NSP299.
FT CHAIN 300 ? NONSTRUCTURAL PROTEIN NSP300.
FT CHAIN 301 ? NONSTRUCTURAL PROTEIN NSP301.
FT CHAIN 302 ? NONSTRUCTURAL PROTEIN NSP302.
FT CHAIN 303 ? NONSTRUCTURAL PROTEIN NSP303.
FT CHAIN 304 ? NONSTRUCTURAL PROTEIN NSP304.
FT CHAIN 305 ? NONSTRUCTURAL PROTEIN NSP305.
FT CHAIN 306 ? NONSTRUCTURAL PROTEIN NSP306.
FT CHAIN 307 ? NONSTRUCTURAL PROTEIN NSP307.
FT CHAIN 308 ? NONSTRUCTURAL PROTEIN NSP308.
FT CHAIN 309 ? NONSTRUCTURAL PROTEIN NSP309.
FT CHAIN 310 ? NONSTRUCTURAL PROTEIN NSP310.
FT CHAIN 311 ? NONSTRUCTURAL PROTEIN NSP311.
FT CHAIN 312 ? NONSTRUCTURAL PROTEIN NSP312.
FT CHAIN 313 ? NONSTRUCTURAL PROTEIN NSP313.
FT CHAIN 314 ? NONSTRUCTURAL PROTEIN NSP314.
FT CHAIN 315 ? NONSTRUCTURAL PROTEIN NSP315.
FT CHAIN 316 ? NONSTRUCTURAL PROTEIN NSP316.
FT CHAIN 317 ? NONSTRUCTURAL PROTEIN NSP317.
FT CHAIN 318 ? NONSTRUCTURAL PROTEIN NSP318.
FT CHAIN 319 ? NONSTRUCTURAL PROTEIN NSP319.
FT CHAIN 320 ? NONSTRUCTURAL PROTEIN NSP320.
FT CHAIN 321 ? NONSTRUCTURAL PROTEIN NSP321.
FT CHAIN 322 ? NONSTRUCTURAL PROTEIN NSP322.
FT CHAIN 323 ? NONSTRUCTURAL PROTEIN NSP323.
FT CHAIN 324 ? NONSTRUCTURAL PROTEIN NSP324.
FT CHAIN 325 ? NONSTRUCTURAL PROTEIN NSP325.
FT CHAIN 326 ? NONSTRUCTURAL PROTEIN NSP326.
FT CHAIN 327 ? NONSTRUCTURAL PROTEIN NSP327.
FT CHAIN 328 ? NONSTRUCTURAL PROTEIN NSP328.
FT CHAIN 329 ? NONSTRUCTURAL PROTEIN NSP329.
FT CHAIN 330 ? NONSTRUCTURAL PROTEIN NSP330.
FT CHAIN 331 ? NONSTRUCTURAL PROTEIN NSP331.
FT CHAIN 332 ? NONSTRUCTURAL PROTEIN NSP332.
FT CHAIN 333 ? NONSTRUCTURAL PROTEIN NSP333.
FT CHAIN 334 ? NONSTRUCTURAL PROTEIN NSP334.
FT CHAIN 335 ? NONSTRUCTURAL PROTEIN NSP335.
FT CHAIN 336 ? NONSTRUCTURAL PROTEIN NSP336.
FT CHAIN 337 ? NONSTRUCTURAL PROTEIN NSP337.
FT CHAIN 338 ? NONSTRUCTURAL PROTEIN NSP338.
FT CHAIN 339 ? NONSTRUCTURAL PROTEIN NSP339.
FT CHAIN 340 ? NONSTRUCTURAL PROTEIN NSP340.
FT CHAIN 341 ? NONSTRUCTURAL PROTEIN NSP341.
FT CHAIN 342 ? NONSTRUCTURAL PROTEIN NSP342.
FT CHAIN 343 ? NONSTRUCTURAL PROTEIN NSP343.
FT CHAIN 344 ? NONSTRUCTURAL PROTEIN NSP344.
FT CHAIN 345 ? NONSTRUCTURAL PROTEIN NSP345.
FT CHAIN 346 ? NONSTRUCTURAL PROTEIN NSP346.
FT CHAIN 347 ? NONSTRUCTURAL PROTEIN NSP347.
FT CHAIN 348 ? NONSTRUCTURAL PROTEIN NSP348.
FT CHAIN 349 ? NONSTRUCTURAL PROTEIN NSP349.
FT CHAIN 350 ? NONSTRUCTURAL PROTEIN NSP350.
FT CHAIN 351 ? NONSTRUCTURAL PROTEIN NSP351.
FT CHAIN 352 ? NONSTRUCTURAL PROTEIN NSP352.
FT CHAIN 353 ? NONSTRUCTURAL PROTEIN NSP353.
FT CHAIN 354 ? NONSTRUCTURAL PROTEIN NSP354.
FT CHAIN 355 ? NONSTRUCTURAL PROTEIN NSP355.
FT CHAIN 356 ? NONSTRUCTURAL PROTEIN NSP356.
FT CHAIN 357 ? NONSTRUCTURAL PROTEIN NSP357.
FT CHAIN 358 ? NONSTRUCTURAL PROTEIN NSP358.
FT CHAIN 359 ? NONSTRUCTURAL PROTEIN NSP359.
FT CHAIN 360 ? NONSTRUCTURAL PROTEIN NSP360.
FT CHAIN 361 ? NONSTRUCTURAL PROTEIN NSP361.
FT CHAIN 362 ? NONSTRUCTURAL PROTEIN NSP362.
FT CHAIN 363 ? NONSTRUCTURAL PROTEIN NSP363.
FT CHAIN 364 ? NONSTRUCTURAL PROTEIN NSP364.
FT CHAIN 365 ? NONSTRUCTURAL PROTEIN NSP365.
FT CHAIN 366 ? NONSTRUCTURAL PROTEIN NSP366.
FT CHAIN 367 ? NONSTRUCTURAL PROTEIN NSP367.
FT CHAIN 368 ? NONSTRUCTURAL PROTEIN NSP368.
FT CHAIN 369 ? NONSTRUCTURAL PROTEIN NSP369.
FT CHAIN 370 ? NONSTRUCTURAL PROTEIN NSP370.
FT CHAIN 371 ? NONSTRUCTURAL PROTEIN NSP371.
FT CHAIN 372 ? NONSTRUCTURAL PROTEIN NSP372.
FT CHAIN 373 ? NONSTRUCTURAL PROTEIN NSP373.
FT CHAIN 374 ? NONSTRUCTURAL PROTEIN NSP374.
FT CHAIN 375 ? NONSTRUCTURAL PROTEIN NSP375.
FT CHAIN 376 ? NONSTRUCTURAL PROTEIN NSP376.
FT CHAIN 377 ? NONSTRUCTURAL PROTEIN NSP377.
FT CHAIN 378 ? NONSTRUCTURAL PROTEIN NSP378.
FT CHAIN 379 ? NONSTRUCTURAL PROTEIN NSP379.
FT CHAIN 380 ? NONSTRUCTURAL PROTEIN NSP380.
FT CHAIN 381 ? NONSTRUCTURAL PROTEIN NSP381.
FT CHAIN 382 ? NONSTRUCTURAL PROTEIN NSP382.
FT CHAIN 383 ? NONSTRUCTURAL PROTEIN NSP383.
FT CHAIN 384 ? NONSTRUCTURAL PROTEIN NSP384.
FT CHAIN 385 ? NONSTRUCTURAL PROTEIN NSP385.
FT CHAIN 386 ? NONSTRUCTURAL PROTEIN NSP386.
FT CHAIN 387 ? NONSTRUCTURAL PROTEIN NSP387.
FT CHAIN 388 ? NONSTRUCTURAL PROTEIN NSP388.
FT CHAIN 389 ? NONSTRUCTURAL PROTEIN NSP389.
FT CHAIN 390 ? NONSTRUCTURAL PROTEIN NSP390.
FT CHAIN 391 ? NONSTRUCTURAL PROTEIN NSP391.
FT CHAIN 392 ? NONSTRUCTURAL PROTEIN NSP392.
FT CHAIN 393 ? NONSTRUCTURAL PROTEIN NSP393.
FT CHAIN 394 ? NONSTRUCTURAL PROTEIN NSP394.
FT CHAIN 395 ? NONSTRUCTURAL PROTEIN NSP395.
FT CHAIN 396 ? NONSTRUCTURAL PROTEIN NSP396.
FT CHAIN 397 ? NONSTRUCTURAL PROTEIN NSP397.
FT CHAIN 398 ? NONSTRUCTURAL PROTEIN NSP398.
FT CHAIN 399 ? NONSTRUCTURAL PROTEIN NSP399.
FT CHAIN 400 ? NONSTRUCTURAL PROTEIN NSP400.
FT CHAIN 401 ? NONSTRUCTURAL PROTEIN NSP401.
FT CHAIN 402 ? NONSTRUCTURAL PROTEIN NSP402.
FT CHAIN 403 ? NONSTRUCTURAL PROTEIN NSP403.
FT CHAIN 404 ? NONSTRUCTURAL PROTEIN NSP404.
FT CHAIN 405 ? NONSTRUCTURAL PROTEIN NSP405.
FT CHAIN 406 ? NONSTRUCTURAL PROTEIN NSP406.
FT CHAIN 407 ? NONSTRUCTURAL PROTEIN NSP407.
FT CHAIN 408 ? NONSTRUCTURAL PROTEIN NSP408.
FT CHAIN 409 ? NONSTRUCTURAL PROTEIN NSP409.
FT CHAIN 410 ? NONSTRUCTURAL PROTEIN NSP410.
FT CHAIN 411 ? NONSTRUCTURAL PROTEIN NSP411.
FT CHAIN 412 ? NONSTRUCTURAL PROTEIN NSP412.
FT CHAIN 413 ? NONSTRUCTURAL PROTEIN NSP413.
FT CHAIN 414 ? NONSTRUCTURAL PROTEIN NSP414.
FT CHAIN 415 ? NONSTRUCTURAL PROTEIN NSP415.
FT CHAIN 416 ? NONSTRUCTURAL PROTEIN NSP416.
FT CHAIN 417 ? NONSTRUCTURAL PROTEIN NSP417.
FT CHAIN 418 ? NONSTRUCTURAL PROTEIN NSP418.
FT CHAIN 419 ? NONSTRUCTURAL PROTEIN NSP419.
FT CHAIN 420 ? NONSTRUCTURAL PROTEIN NSP420.
FT CHAIN 421 ? NONSTRUCTURAL PROTEIN NSP421.
FT CHAIN 422 ? NONSTRUCTURAL PROTEIN NSP422.
FT CHAIN 423 ? NONSTRUCTURAL PROTEIN NSP423.
FT CHAIN 424 ? NONSTRUCTURAL PROTEIN NSP424.
FT CHAIN 425 ? NONSTRUCTURAL PROTEIN NSP425.
FT CHAIN 426 ? NONSTRUCTURAL PROTEIN NSP426.
FT CHAIN 427 ? NONSTRUCTURAL PROTEIN NSP427.
FT CHAIN 428 ? NONSTRUCTURAL PROTEIN NSP428.
FT CHAIN 429 ? NONSTRUCTURAL PROTEIN NSP429.
FT CHAIN 430 ? NONSTRUCTURAL PROTEIN NSP430.
FT CHAIN 431 ? NONSTRUCTURAL PROTEIN NSP431.
FT CHAIN 432 ? NONSTRUCTURAL PROTEIN NSP432.
FT CHAIN 433 ? NONSTRUCTURAL PROTEIN NSP433.
FT CHAIN 434 ? NONSTRUCTURAL PROTEIN NSP434.
FT CHAIN 435 ? NONSTRUCTURAL PROTEIN NSP435.
FT CHAIN 436 ? NONSTRUCTURAL PROTEIN NSP436.
FT CHAIN 437 ? NONSTRUCTURAL PROTEIN NSP437.
FT CHAIN 438 ? NONSTRUCTURAL PROTEIN NSP438.
FT CHAIN 439 ? NONSTRUCTURAL PROTEIN NSP439.
FT CHAIN 440 ? NONSTRUCTURAL PROTEIN NSP440.
FT CHAIN 441 ? NONSTRUCTURAL PROTEIN NSP441.
FT CHAIN 442 ? NONSTRUCTURAL PROTEIN NSP442.
FT CHAIN 443 ? NONSTRUCTURAL PROTEIN NSP443.
FT CHAIN 444 ? NONSTRUCTURAL PROTEIN NSP444.
FT CHAIN 445 ? NONSTRUCTURAL PROTEIN NSP445.
FT CHAIN 446 ? NONSTRUCTURAL PROTEIN NSP446.
FT CHAIN 447 ? NONSTRUCTURAL PROTEIN NSP447.
FT CHAIN 448 ? NONSTRUCTURAL PROTEIN NSP448.
FT CHAIN 449 ? NONSTRUCTURAL PROTEIN NSP449.
FT CHAIN 450 ? NONSTRUCTURAL PROTEIN NSP450.
FT CHAIN 451 ? NONSTRUCTURAL PROTEIN NSP451.
FT CHAIN 452 ? NONSTRUCTURAL PROTEIN NSP452.
FT CHAIN 453 ? NONSTRUCTURAL PROTEIN NSP453.
FT CHAIN 454 ? NONSTRUCTURAL PROTEIN NSP454.
FT CHAIN 455 ? NONSTRUCTURAL PROTEIN NSP455.
FT CHAIN 456 ? NONSTRUCTURAL PROTEIN NSP456.
FT CHAIN 457 ? NONSTRUCTURAL PROTEIN NSP457.
FT CHAIN 458 ? NONSTRUCTURAL PROTEIN NSP458.
FT CHAIN 459 ? NONSTRUCTURAL PROTEIN NSP459.
FT CHAIN 460 ? NONSTRUCTURAL PROTEIN NSP460.
FT CHAIN 461 ? NONSTRUCTURAL PROTEIN NSP461.
FT CHAIN 462 ? NONSTRUCTURAL PROTEIN NSP462.
FT CHAIN 463 ? NONSTRUCTURAL PROTEIN NSP463.
FT CHAIN 464 ? NONSTRUCTURAL PROTEIN NSP464.
FT CHAIN 465 ? NONSTRUCTURAL PROTEIN NSP465.
FT CHAIN 466 ? NONSTRUCTURAL PROTEIN NSP466.
FT CHAIN 467 ? NONSTRUCTURAL PROTEIN NSP467.
FT CHAIN 468 ? NONSTRUCTURAL PROTEIN NSP468.
FT CHAIN 469 ? NONSTRUCTURAL PROTEIN NSP469.
FT CHAIN 470 ? NONSTRUCTURAL PROTEIN NSP470.
FT CHAIN 471 ? NONSTRUCTURAL PROTEIN NSP471.
FT CHAIN 472 ? NONSTRUCTURAL PROTEIN NSP472.
FT CHAIN 473 ? NONSTRUCTURAL PROTEIN NSP473.
FT CHAIN 474 ? NONSTRUCTURAL PROTEIN NSP474.
FT CHAIN 475 ? NONSTRUCTURAL PROTEIN NSP475.
FT CHAIN 476 ? NONSTRUCTURAL PROTEIN NSP476.
FT CHAIN 477 ? NONSTRUCTURAL PROTEIN NSP477.
FT CHAIN 478 ? NONSTRUCTURAL PROTEIN NSP478.
FT CHAIN 479 ? NONSTRUCTURAL PROTEIN NSP479.
FT CHAIN 480 ? NONSTRUCTURAL PROTEIN NSP480.
FT CHAIN 481 ? NONSTRUCTURAL PROTEIN NSP481.
FT CHAIN 482 ? NONSTRUCTURAL PROTEIN NSP4
```

Db 40 VTTAAQRAIV-----AVIPRVEFOMQVS-----DHP---AL 69
QY 63 HNELEQCRARAGRCLEVGAMHPS-----INDNPVLHRCLEIRPVGDVOMYS----- 111
Db 70 H-AISRTTR--HWIEMG--PKKALHVLIDPSGL-----LRVAVAEERWALCLHRT 118
QY 112 -----APTRGRAN-----CRSALRGLP-----ADRTYC-FDGF 141
Db 119 ARKATATALAEASAMHADVCC--ALRGAPSGFVYHVEDVPHGRVADRCLLTYTPM 175
QY 142 SRCFAAETGVALXSLHDLMPADVA-----E 167
Db 176 QKCELMATITATLLVAVDLMPVLAIAHVGDMDLGTAMHLHDGCCPADRCAGAPTP 235
QY 168 AMARHGATRLAALHLRPEVLLRPGYHTTSYLLIHGDRAVVTYEGDTSAGYHNDVSL 227
Db 236 GYTRPCTRTIYQVL--PDFAHGRGLYRCGRPLMTRDCAVAELSWEVAQHCQH----- 286
QY 228 RANITRTKIYGDHPLVIERVRAIGCHEVLLTAAPSPMPYVYPRSTEVYVRSITPG 287
Db 287 -----ARVAVRCTLPIRHVSLOPS-----ARVR----- 311
QY 288 GSPSLPSSACSTSTFAVYVHIMDLRMLFGATLLDQAFCCSRMLMTYLRGISYKVTYAL 347
Db 312 -LPDLVHLAEGWRFSLRPVFORMLSTCKTSLSPDAYISERFEKKNALCHSITLAGN 370
QY 348 VANEGNN--ASEDALTAITAAVLTICHOVYRTQAISKMRRLGVEHAKOFTTRLS- 403
Db 371 VLOEGMKGTCABEDALCA-----YAFRAMQSNARLAGIMKAKCAADSLV 417
QY 404 --WLFESGSDIYIPGROLOFYAOCRRWLSA-----GHLDPRVLVFESYPCRCRTF 453
Db 418 AGML-----DTI-----WDAIKRFLGSVPLAERMEEMODAAVAAPF----- 454
QY 454 LKKYAGFCFCFMWMLGOECTFLPAPAGLVGDHGDHNEAVEGSEVDAEPAHLDVSTYA 513
Db 455 -----RGLPDEGRIHLDIYO-----PKSPREPIATWTI 484
QY 514 VHGQLE--ALYRALNVPD-----IAARAS 537
Db 485 VHAASEDRHCACAPRCVPREREPAPAGOPDDEALLRPMLEAERALLRCREMFEALRAR 544
QY 538 RLATATVELVASPRLBECTVL-----GNKTFRTYVDGAHLEANGPEYOY 581
Db 545 ADTAAAPAPAPAPAPARYPTVLYRHPAHNGPWLTLDEREADALVLDDPLGOLRGERH 604
QY 582 VLSEFDSRQSMGAGSHSLTGLTLPAGLO--VRISNGLDCTATFPPGAGSAAPGEVAAP 639
Db 605 -----FAAGAHMCA--QARGLOAFVR-----VPPERPMWADGG--ARA 639
QY 640 CSALYRYNRFTQSHSLTGLMLHREGLG---IFPPESGHIMESANPFCGEGTLYTRT 695
Db 640 WAFFPFGCAAAQR-----LLGERPAVMHLRYTGDVDPOLI-----ALALRT 679
QY 696 WTSFGSSDS-----PREAAPAMAAT-----PGLP-----HS 724
Db 680 LAQOGAALASVDLPEGGAFFDANAATAAVRAGRQSAASPPGDDPPRRARRSORHS 739
QY 725 -----TPVSDIWLPPSEEFQVDAAPVPAPAPAGLPGFVLTLPVPPVHK----- 773
Db 740 DANGTTPPAPARDPPP-----APSPAPPRAGDPV-----PIPAGADARDAELE 787
QY 774 -----PSIPEPSSNR-----RLTYYPDGAKVYAGSLFESDCML 808
Db 788 VACEPSGPSTRADPDSDIVESYARAAGPVHLRVDIRMPGCKV-----V 835
QY 809 VNASNGHRPBGGLCHAFYORPEAPYPTFEIMREGIAA--YTLTPRP----- 854
Db 836 VNANEGGLAGSGVCAIFENATAA-----LNAACRLAPDPGGEAATPGHGC 884
QY 855 -----IHAAP-----DYREONPKRLAAYRE-----TCSRGTAYAPLLGSGIY--QY 898
Db 885 GYTHIIHAAPRRPRPALLEEGBALLERAYRISVALAARRMACVACPLGAGVTGWSA 944

QY 899 PVLSSEDAWERNHRPGDELYL-----TEPANWFE 928
Db 945 AESLRALAAATPREPERVSVLHCHPDRAITLTHASVLYGAGLAARRVSPPTPELAS-CP 1003
QY 929 ANKPAQVLTITTEDTARTANLALEIDAATEVGRACACTIS----- 969
Db 1004 AGDPRAPRASAPPAPIG-----DATAPEPRGCGCELCRYTRYVNDRAVYVNLWLERD 1058
QY 970 -----PGIV-----HYOF----- 977
Db 1059 RGATSWAMRIPEVYVYVPELATHFPLNHTSVLKPAEVPRPGMGSDMRGCMHGMPO 1118
QY 978 -----TAGVPSGKSRS10QGDVYVVPTRRLRN--SWRRGFAATPTHTAA 1023
Db 1119 VRCTPSNAHALCRTGVPRPASTRG--GELDPNTCWLRAANVAQAAACAGAYTSAGCP 1175
QY 1024 RVITGR-----RVVIDEAP 1037
Db 1176 KCAVGRALSEARTHEDEFAALSORMSASHADASPDGTGDLPLMETVGCASRWVWSEH 1235
QY 1038 SLPPHLLHMOA-----SSVHL-----GDPNQI--PAI----- 1066
Db 1236 EAPPDHLVSLHRAPNCPMGVYEVNARPEGPNPTGHFVAVAGGPRRYSDRHMLAVP 1295
QY 1067 -----DPE-----HAGLYPAIRPELAPS--WMXYT-- 1090
Db 1296 LSRGGTCANTDEGLAOAYVDLEVRRLDGDAMARALASVQRPBKOPYNIRVWNMAGA 1355
QY 1091 -----HRCPA-----DYCELIR-----GAYRKI-- 1108
Db 1356 GKTTRLAATREDLVYCPNALLHETIQAKLRADIDIKNAATYERRLTKPLAAYRITYI 1415
QY 1109 -----QTSRYL-----RSLF--WNEP-- 1123
Db 1416 DEAFITLGECECAVAAQTTHAEVLCVGDROCGPHYANNCTPPYDPMPTFRSHTTRFPD 1475
QY 1124 -----AIGOKLYXTOAKANPAGAITVHE 1147
Db 1476 CWAARLRAGLDYDIEBERGTFCACNLMDGRQVDLHLAFSRETV--RRLHAGIRATVRE 1533
QY 1148 AGCATFTEETIIMTADARGL--IOSRAHVALTRHTEKEVILD--APGLREYIGSDVY 1205
Db 1534 AQCMSVGTACIHVGRDGTVALTRDLATVSLTRASDALIYHELEDGSIKRAAGLS----- 1589
QY 1206 NNFFLAGGEV-----GXNR-----PSVIPRGNPDONLCTLOAFPPSCQISAYHOL 1250
Db 1590 --AFDAGALAELEKEVPAGIDRVVAEQAPRPLP--PADGIEADQDVPPFCRTELELV 1644
QY 1251 ABEIGHRPAPVAAVLPPCELEOGLYMPQELTVSDSVLFELTDIVHCHMAAPSOAKAY 1310
Db 1645 FGRAGH--PHYADLNKRVTEGEKEVRYM-----RISRHLKMKNHTEM-----PSTER-V 1689
QY 1311 LSTLVGRYGRRTKLYEAAH--SDVRESLARFP-----TIGPVQATV-----CELYEL 1356
Db 1690 LSAVCA-----VRRYRAGEGSLTRTAVANORHPRPRQIPLPPRYTAVGAOEMRATYLRER 1744
QY 1357 VEAAMVERGODGSAVLEL-----DLGNRVSRITTFQKXCNKR-----T 1394
Db 1745 IDLTIDVYTGQVGAARELTDIBYARPEEIFAGMCTAGSLSPATLAKTKVDADALPPOI 1804
QY 1395 TGETIAHGKVGCGISAMSKTFCALFGWFAIKEILLALPPIFYGDAYEESVFAAAYS 1454
Db 1805 EDCHAQKAGLEIRAMAKEMWVQVMSPHFRAIOKIIMRALRPQFLVAAGHTEPEVDAMMO 1864
QY 1455 GAGSCAVFENDPSEPDSTONFS-----LLECYVMEECGMPOMLILYTHVLS 1503
Db 1865 AHYTTNAIEVDTEFPMNQTALATRDVELEISAALLDPCA--ED-----YALRA 1912
QY 1504 -AMILQAPRESLKGFKKHSGEGTLLIMNTVMMAIICACEYR-----DFRYVA--FKGDD 1557
Db 1913 GSYCTIRELGSTETGECERTSGEPATILLHNTT-----VAMCMAMRMVPKGVNAGITOGDD 1967


```

OY 1138 -----ANPGL-TVEHAGATFETTTI-ATADAGLISSRAHAIVALTIRHTEK- 1185
DB 1018 LKRGEEVSTFEINIVHIOGETFEDVSVRLPTPELISSSPHVIALVLRHRTKSE 1077
OY 1186 ----CVLDAGGLREV-----GISDIVNNFPLAGSEVGHRSVIRPGNPDNLGLQA 1237
DB 1078 KYYSVADLP---LVKWSDSLKVSDFLDMKVDAGLIXLOLQSGSTKRG--ENL----- 1126
OY 1238 FPSQCSAISAHQIAELGHRPAP-VAAVLPPCELEOGLLYMPQELTVSDSVLEFELTDI 1296
DB 1127 FVP-CPRSGYISDMQYVDFLPGNSFILNEDYAVTNL--RENNLVKCTCIDFSKS-- 1181
OY 1297 VRCMAAPSORKAVLSTLVRCYGRKRLYCAASDVRESIARLPV--IGPVQATTCELY 1354
DB 1182 ----VSVPRQOGEFFP-----AHRTAERPRRSAGILENIVAMIKRNFNSPDLTGLIDIE 1232
OY 1355 ELVEAWEKQDSAVLELLCNRDVSRIFFOKXC--NKFTTGENI----- 1399
DB 1233 DTALVYNNKWDVYIIDLSCGNVTPMTSDAFHMAKOEKSTIGOLADPDVLDPAIDQ 1292
OY 1400 -----AHGRVGGGISA-----WSKTEFCALFGIWFRAIEKEILLALPPN--IF 1439
DB 1293 YKHMINKAPQOKIGLSPQDEYALQTIIVHSKOINAFGLFSELTRQLERIDSSKFLF 1352
OY 1440 YGAAEESV--FAAASGASCSWFENDESEPDSTONNESLGLCEVYMESGMPWLIRL 1497
DB 1353 YTRKTPQIAEAFSDSDSTVPMVELIDISKYPSQNEFCAVEYLIMELKINGFL-- 1409
OY 1498 YHLVRSAMILQAPRESLKGK-----WKHSGEPTLLMNTVMMAIIAHACEFRDF 1548
DB 1410 ----EEVWKQGHKRTSLKDYTAGIKTCLMYOKRSGDVTITGNTVITIAACLASIIPMDKV 1465
OY 1549 RVAAFKGDSDSVLCS--DYRQSRMAALJAGGLKLVDRP-IGLYAG-VVYAPGLGT 1603
DB 1466 IKAFCGDDSDILYIPKGLDLPDIOSGANLWMNEFAKL--YRKRYGFCRYIIHHDRGA 1522
OY 1604 LP--DVVRFGRLESEKMGWGPERRAEQLAVCDPLGLGNVAGV--CVVVSRYVYVSP 1659
DB 1523 IYVYDPLKLKLGCKHI-KSLDLEEFRLSLDVSLSLNCAYFGOLDALAEVH---- 1577
OY 1660 GLVHNLGMLQTLADGKAHFTETIKPYLD 1688
DB 1578 -----KTAVNGSFAFCSTVKYLSLD 1596

RESULT 8
POLR.TYMW STANDARD: PRT: 1844 AA.
AC F10358;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA REPLICASE POLYPROTEIN (EC 2.7.7.48).
OS Turnip yellow mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.
OX NCBI_TaxID=12154;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88289359; PubMed=3399388;
RA Morch M.D., Boyer J.C., Haenni A.L.;
RT "Overlapping open reading frames revealed by complete nucleotide
RT sequencing of turnip yellow mosaic virus genomic RNA.";
RL Nucleic Acids Res. 16:6157-6173(1988).
CC -I- MISCELLANEOUS: THE 206 KDA PROTEIN IS POTENTIALLY A POLYPROTEIN
CC (BY HOMOLOGY TO THE LONGER PROTEIN OF THE SINDIS VIRUS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: X07441: CAA30322.1; ALT_SEQ.
DR PIR: S01956; S01956.
DR MEROPS: C21.001; -.
DR InterPro: IPR006066; -.
DR Pfam: PF01443; Viral_helicase1, 1.
KW Transferase; RNA-directed RNA polymerase; Polypotein; ATP-binding.
FT NP_BIND 976 983 ATP (BY SIMILARITY).
FT DOMAIN 1631 1665 POLYMERASE SITE (BY SIMILARITY).
SQ SEQUENCE 1844 AA; 206640 MW; A016D758C83D128C CRC64;

Query Match 3.9%; Score 349; DB 1; Length 1844;
Best Local Similarity 20.6%; Pred. No. 3.6e-14;
Matches 385; Conservative 203; Mismatches 649; Indels 634; Gaps 93;

OY 157 LHD-----LMPADVAEMARH-GXTRLYAALHLPPEVLPRG-----TYHTSYLLIHD 204
DB 141 MHDALMYHPSQIMDLFLKRPNIERLYASLVPEPAHLSDQSFYPKLYTYTTRHLLHY- 139
OY 205 GDAVVYIEGDTAGYNHDSILRAMITRKI-VGDHPLIEFVRAIG-CHEVILLTAAP 262
DB 200 -----VPEGHEAGSYNP-SDAHSMIRISILGNHLSVTLLSESGPVHSLIGRGP 252
OY 263 EESPMYVPRPRSTEVYVNSIFGP----- 286
DB 253 PPDSIQAPPTLMTSDLFRSYQEPRLDVVSFRIPDAIELPQATFLOQPLRDLVPRAVYN 312
OY 287 -----GGSPLFSPSACSTKSTFNHAYVHIMDLMLFGATLDQACCSGLMT 333
DB 313 ALFTYRAVNTLFTSDPAALVRHSSKPDHDWTSNAMDVLQTF-----ALLNV 361
OY 334 YLR-GISYKVTVG-----ALVANEGNWASEDALTAITAAYLTICQRYLRTQASIKMR 387
DB 362 PLRPNVYVHQLSPIASLSLYLRQHRRL--TATAVPLISLELL--QRL----- 408
OY 388 RLCEVHAQ-KFTIRLYSWLEFKSGRDYIPGRQLOFAQCRMLSAGFHLDPRVLVDESY 446
DB 409 PLPIPLAEVKSITAFRELYRKKE-----PHHPLDV-----FHLOHRYVNHSAI 453
OY 447 PCGRFFLKVKVAGKPCCFMMLQOECTCLFEPABGLVGDHNDNEAVEGSEVDPAPAH 506
DB 454 -----SAVRPASPPHQ 464
OY 507 DVSGTYAVHGHOLEALYRALNVPDIAARA-----SLRTATYELVASPPR--- 551
DB 465 KLP-----HALOKAALLLRIPSLTATPFFRSEOKSMPLNALSMTLKRPLMPQASL 519
OY 552 ----LECRVYLGKRTFRTYVVDGAHLEANGPEQYVLSFPAASROSGAGSHSLTYELTPAG 607
DB 520 VLLALSESSITLHKRLSPPTLQAH-----DTYHRLHPGSYSIQWERTP-- 564
OY 608 LQYRISNGL--DCTATPPGAPSAAGEVAAFCSALYRYRFTQRIHSLTGGLMLPE 664
DB 565 LSLPRTTALPLPTPTTSTAPPDRSEASLP--PAFAST-----FVPR----- 603
OY 665 GLLGIRPP--FSPG--HIMESANPECGGLTYRTWSTSGF-----S 702
DB 604 -----PPRAASSPGAOPRTTAAPTPIEPT--QRTHNSDLALSSSTTEPPPPINS 655
OY 703 SDFSP-----PEAAAPMAATPGL--PHSTPPVSDIWLPPPEE----- 740
DB 656 PDMTSPAPVLPFETNSPPRRPPQLPATRPDLERAHMPPLSIPHQDPTDSADPLMGSHLLH 715
OY 741 FOYDAPVPPAPPAOLPGPVYLTPPPPPV-----HKPS-- 775
DB 716 HSLPAPPTPLPSQLLPAPLTNDPPAIGVLPFEELHRRYPENTATELTRLSLPSSNH 775
OY 776 IPPSPNRRLLYTPPGAVYAGSLFESQCDMLVNA--SNPGHPPG-----GLCH-- 824
DB 776 LPOPTLN-CLLSAVSQOTVSEHLMWESIQTILPDSQLSNEENTNIGLSTENHTALAHLY 834
OY 825 -----AFYORFPEAFVTEFIMREGIAAYTLPRPIIHAVAPDYRVEONPKRLEAVRET 879

```



```

Db 835 NFOATVYSDBRGLLPSPDITK-----IDTHTTGPSSHRSPG-KRLLSGP-- 861
QY 880 CSRRGTAAYPLLG-----SGIYQVPSLSPDAMERNHRPGDELY---LTPRANMFEA 929
Db 882 -SAKQHPSPDLIRAMKSPFKSGNY-LPFS-----EAHNPTSISIAKNLISMKNQFPG 933
QY 930 NKAQAVLTITETARTANALEIDAETVGRACAGCTISPG-----IVHQFATAGVPGSG 965
Db 934 VLSLSTVSTGQRTGPPRKERTIQIDHYLDT-----NPGKTPVYHF--AGFAGCG 981
QY 986 KSRISIQ-----GDVYVYVPTRELNSMR--RGFAA--FTPHATVYTGIRGV 1032
Db 982 KTYPIQGLKTKIKFKRFR-VSCPTTELKTEKTKAMELHGSQSRFTWESSILKSRILV 1040
QY 1033 IDEAPSLPRLHLLHMOARSSVHL--LQDP-----NQIPAIDEHAGLVP 1075
Db 1041 IDEIYKMPKGYLDLSILADPALELYILGDPLOGEVHSGSKSDSNRRLPS--ETLRLLP 1097
QY 1076 AIRPELAPTSMMXVTHRCRADVCEL--RGATYRIQT-----TSRVLRSLF 1119
Db 1098 YI---DMTCWM--SYRIQCIARLFOIHSFNAMQVIGSVSTPHQSPVLTKSHASSLT 1151
QY 1120 WNEPAIGOKLVYTOAKAANPGATVHEAOGATFETITATADARGLIQSSRAHAIVAL 1179
Db 1152 FN--SLGFR-----SCTISSQGLFCQDPATIVDNYTKWLSS--ANGILVAL 1194
QY 1180 TR-----HTEKCVILDA--PGLREVGISDVIYNNFLL 1210
Db 1195 TRRSRGVQEMGPFSSVYGGTNGSSAMFSDAFNNSLIIMDRFPLFPOL--KLITSPLTT 1251
QY 1211 AGGEVGHKRP-----VIRGPNQDNLGLQAPPS 1241
Db 1252 RGKLLGATPSPASPTHRSPFHLPHIPLSYDRDEVTVNFTLPDQGPERRLOT--HFLRP 1309
QY 1242 COISAFHOLAEELGHRPAVAVALPCCPELEOGLILYMPOLFTVSDSLVLEL-----T 1294
Db 1310 SRLPLHFDLPRAI--TPPYVSTVDP--POAKAPVY--PGEFF--FDLSIAFELPAHDPSTR 1363
QY 1295 DIVHCHMAAPS---ORKAVLSTLVGRYGRRTKLYEAAH-----SDVRESLARFT 1340
Db 1364 EIIHKQSSNOEPMFDRPFSLSG-----QSSLLISAKHAPNHDPTLLPASINKRL--RFR 1416
QY 1341 PTIGPVQATP-----C-----ELYELVEMVKGQDGSVALE 1372
Db 1417 PDSQPHQITADADVLTGLQFLHSLCRAYSROPNSTVPPNPDELFAECISLNEYAQLSSKTOS 1476
QY 1373 LDCLNBDVS-----RITFEQKXCKNFTTGETIANGKVGQGISAMSKTECALFGP--W 1422
Db 1477 TIYANASRSDPWRHTTIVAFKAQHKVNDGSTIFGSMKACQITLALMHDYIYLVGPKKY 1536
QY 1423 FRAIEKEILLALPPNIF-----YGDAYEESVFAAASGAGSCWVFENDESEDDSTQ 1473
Db 1537 ORTFEDN---ADRPNIYSHCGKTPNOLRDMCOEHL-----THSTPIKLANDYAFDQSO 1586
QY 1474 NNSLGLCECVAMECGMPQLRLY--HIVRSANILLQAPKESLGFKKHSGEGTLLMNT 1532
Db 1587 HGESVYLEAKMKRLNIPSLIQLHVLKTNVSTQGFPLTCM-----RLTGEGETYDDNT 1641
QY 1533 VMMMAITACYERFDRVAFAFGKDSVVLCSDY--RQSRNAALLIAGCGKLVDY---- 1586
Db 1642 DYVLAVIYSGYDVGSCPRAW--SGDDSLI---DHPLETRHDPVSVLRLHRLFELELSHP 1697
QY 1587 -----RPIGLYAGVVAAGLGLTLPD---VVRF--AGRLSEKKNWGPQPER 1625
Db 1698 LFCGYVVGAGCIRNPLALFCKIMIAVDDADLDRSLYTEFTTGHLLGESIMHLLPET 1757
QY 1626 ABOLRLAYCDF 1636
Db 1758 HVOYQACEDF 1768

```

RESULT 9

```

POLR_TYMC
ID POLR_TYMC STANDARD: PRT; 1844 AA.
AC P28477;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE RNA REPLICASE POLYPROTEIN (EC 2.7.7.48).
OS Turnip yellow mosaic virus (isolate TYMC).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tymovirus.
OX NCBI_Taxid=31751;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92119261; PubMed=1731998;
RA Dreher T.W., Branson K.L.;
RT "Genomic RNA sequence of turnip yellow mosaic virus isolate TYMC, a
RT cDNA-based clone with verified infectivity."
RL Plant Mol. Biol. 18:403-406(1992).
CC -1- MISCELLANEOUS: THE 206 KDA PROTEIN IS POTENTIALLY A POLYPROTEIN
CC (BY HOMOLOG TO THE LONGER PROTEIN OF THE SINDBS VIRUS).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC
CC EMBL; X16378; CAA34415.1; -
CC DR PIR; S19372; S19152.
CC DR InterPro: IPR000606; -
CC DR Pfam: PF01443; Viral_helicase1.1.
CC KW Transferrase; RNA-directed RNA polymerase; polypotein; ATP-binding.
CC FT NE_BIND 976 983 ATP (BY SIMILARITY).
CC FT DOMAIN 1631 1665 POLYMERASE SITE (BY SIMILARITY).
CC SO SEQUENCE 1844 AA; 206612 MW; 02CB928FCCCA5E81 CRC64;

```

Query Match 3.8%; Score 346; DB 1; Length 1844;
 Best Local Similarity 20.5%; Pred. No. 5,6e-14;
 Matches 380; Conservative 210; Mismatches 660; Indels 608; Gaps 94;

```

QY 157 LHD-----LMPADVAEAMARH-GXTRLYAALHLPREVLPPG-----TYHTSYLLIHD 204
Db 141 MHDALMTYHPSQIMDLFLKPNLERLASLVPEEHLSSQSTPKYLTTRHLYL- 199
QY 205 GDAVVYTEGDTAGYNDVSIIRAWITRTKI--VGDPVIEVERAIG--CHEVLLTAAAP 262
Db 200 -----VPEGHAGSYNQ--SDAHSWLRINSIRLGNHLSVTLLSESGPYHSLIQGTP 252
QY 263 EESPMPYVP-----YRST-----EYVNSIFG 285
Db 253 PPDPSIQAPPTLMAASDLFRSYQEPRLDVVSFRIPDALELPQAFLEOQPLRDRVLPRAYVN 312
QY 286 P-----GSPSLPSPASCTKSTFNAPVNHIMDMLMFGATLDQOACCSRLMT 333
Db 313 ALFTYTRAVTLRTSDPAALVRMHSSKPDHDWVTSNMDNLQF-----ALNV 361
QY 334 YLR-GISYKYTVG-----ALVANEGMNASSEDALTAAYTLTICHQRYLRTQAIKSGMR 387
Db 362 PLRPNVYVHYLQSPIASLSLYRQHMRL--TATAVYILSFLLL--QRL- 408
QY 388 RLCVEHNAQ--KETRLYSWLFEEKSGRDYIPGRQLQFYAQCRRMLSAGFNHDPRLVLEDES 446
Db 409 PLPIPLAEVKSITAFRELYRKKE-----RHHPRDV-----FHLOHRIIRNYHSAI 453
QY 447 PCRCRFFLKVKVAKFCFCFMWMLGOEECTFLPRAGLVGDHGNENAESEGVPAPEAH 506
Db 454 -----SAVRPASPHQ 464
QY 507 DVSQTYAVHGHOLEALYRALNVPQDIARA-----SRITATVELVASPDR-- 551
Db 465 KLP-----HALQKAAALLLRIPISPLTATPFFRSEQSKMLPNDELSTWTLKRFPLPQASL 519

```

```

QY 552 ----LECTVGNKTFRTTVVDAHLEANGBPQVYLSFDSARQSGASHLTYELPAG 607
D 520 VLLALSESSILHKLFPPTLOAQH-----DTHRHHLHPGSSYLQWERTP-- 564
QY 608 LQVRSSNGL---DOTAFPPGGAASAPAGVAAFCALYKYNRTOR---HSLTGSLW 660
D 565 LSTPPTTAAPPTTSTAPDRSEASLP---PAFAST-----FVPPPPAASSPGA-- 613
QY 661 LHPBGLIGIFP-PSPGHIMESANFCEGLTYRTWSTGSSDFSP-----P 708
D 614 -QPTTTAAPPPTPIEPPOAHQNSDLALESSTPPTPPPIRSPTMTSAVYLPETNSP 672
QY 709 EAAPAMATPGL--PHSTPPVSDIWLPPSEEFQVD-----AAYPPAPD 753
D 673 RRPPOLPATDLEBAHPPPLS-----IPHODPTSDVPLMGSHLHSLAPPTPLPS 728
QY 754 PAGLPQVYVLPTRPPPV-----HKPS--TPPSRNRLTYT 788
D 729 SOLPAPLTPDPTAIGPLPEELHPRYPENTATFTRLKSLPSNHLPOPTLN-CLLSA 787
QY 789 YPDGAKVYAGSLFESDDCWLVA--SNPGHRPG-----GLCH-----AFQPREPE 832
D 788 VSDOTKXSEELHMSLQITILPDQSLSNETWTGLSTHLLALAHLYFOATVYSDRGPI 847
QY 833 AFYPTFIMREGLAAYTLTPRIIHAVADYVBDQNPRLBAAYRETCSRGTAAVPLG 892
D 848 LFGSDFTKR-----IDITHTGPPSHFSPG-KRLGSGQ---SAKGHPSPDLIR 893
QY 893 -----SGIYQVPSLSFDMERNHRGDELY---LTERPANNFEAKRPQVLTITED 942
D 894 AMKSFVSGN-LPPS-----EAMNHTSISHAKNLSNKNNGDGLSLDVTGORT 946
QY 943 TARTANLALTEIDAETEVRACAGCTISPG---IYHOFTAGVPGSGKRSIQ----- 992
D 947 GPTPKERIQIDHLYDT-----NGKTPPVHF--AGFGCGCTTYIQQLTKKL 994
QY 993 -GDVAVVVPRLRLNSRR---RGFAA--FPTHTAARTVIGRRVVIDEAPLPHLL 1045
D 995 FKDR-VSCPTELTETKKTAMELHGSQSMRFNWESSILKSSRLVDELYKKMRGYLD 1053
QY 1046 LHMGRASSVHL---LGDV-----NOIPADFEHAGIYPAIRELAPTSWMX 1088
D 1054 LSLIADPLLEVIITDGLQGEYHSQKSSNHRPLS---ETLRLPLT---DMYCW- 1105
QY 1089 VTHRCPADVCELI-----RGAYPKIQT-----TSHVLSLFENEPACOKLYXT 1132
D 1106 -SYRLPOCIARLFOIHSFMANQVIGSVTFHDOSPVLTNHASSLTFN--SLGVR-- 1158
QY 1133 QAKAANAGAITVHAQCATETETIATADARGLIOSSRAHAIVALT- 1181
D 1159-----SCTISSQGLTFCDPAIIVLDNYTKWLS--ANGVALTRSRSGVQFMGPS 1207
QY 1182 -----HTEKVIIDA--PGLIREVIGISDVYVNFELAGVGXHRP--- 1220
D 1208 SYVCGTGGSSAMFSDAFNNSLIIMDRYFPLPOL---KLITSPUTTRGRLNGATPSAS 1264
QY 1221 -----SVYPRGNPDONIGTLQAPPPSCQISAYHQLAEEL 1254
D 1265 PTHRSPPHLPRIPLYSYDRDFYVYNSTLIPDQGPETRLDT--HFLPERSRLHLHDLPAI 1322
QY 1255 GHRAPAAVLPPOPELEQGLLYMPQELTVSDVLYVEL-----IDIVYCRMAAS- 1305
D 1323 -TPPVPSTVDP-POAKASPVY-PGEF--PDSLAFFLPADHDESTRILHKDDSSNOFP 1376
QY 1306 -QKAVLSTLVGRYGRRTKLYEAAH-----SDVRSRLARFPTIGPVQAT--- 1350
D 1377 WEDRPFSLSL-----QPSLSLSAKHAPNDPPLLPSINRL-RFRSDSPHITADV 1429
QY 1351 -----C-----ELYELVAMVEKGDGSAVLELIDLCNRDVS----- 1381
D 1430 VLGQLFHLSCRAVSRQPNSTVPPNPELFAFCISLNEYAQLSSTQSTIYVANAQRSDPDW 1489

```

```

QY 1382 ---RITFOKKXCNFTTGETTAHGKVGOGISAMSKTFCALFQ---WEPAIEKELLALP 1435
D 1490 RHTTVKIFAKAOKHRYKNGSIFGSKACOTLALMDYVLLVGPVKYORIPDN---ADRP 1546
QY 1436 PNIF-----YGDVVEESVPAAYAGSACSWFENDESEPTONNPFSLGECVME 1486
D 1547 PNITSHCGKTPTNOLRDMQCEHL-----THSTPKIANDYAFQSOHGESVLEALKMK 1599
QY 1487 ECGMPOWLIRY-HLVRSAMTLOAPKESLKGFKKHSCEPTLLMTVMNNAIIAHCEYF 1545
D 1600 RLNIPIHILQIAHVKHTVSTQFGLTCM-----RLNGEPTDYDNDYINLAVIYQSDV 1654
QY 1546 RQFVAAFKGDSVYVLSQV---KOSRNAALIIACGKLKAVY----- 1586
D 1655 GSCPIMV-SCDSDSLI---DHPPLTRHDMPVYLKRLHLEKLELHSHPLFCGYVYPACGI 1710
QY 1587 -RPIGLVAGVYVAVGLGLTP-----VYRF--AGRLSKNMGPPPERAEQRLAVCF 1636
D 1711 RNPLALFCKLMAIVDDALDRLSYLTMEFTTGHILGSLMHLLEPHVYOQACFDF 1768

```

RESULT 10

RRPO_ORSYVC STANDARD: PRT: 1612 AA.

```

AC P9659: 039640:
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTRAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OC Odontoglossum ringspot virus (Isolate Korean Cy) (ORSV-Cy).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=138661;
RN [1]

```

RP SEQUENCE FROM N.A.

RX MEDLINE=96381046; PubMed=8789059;

RA Ikegami M., Isumura Y., Matsumoto Y., Chataani M., Inoue N.;

RT "The complete nucleotide sequence of odontoglossum ringspot virus (Cy-1 strain) genomic RNA.";

RL Microbiol. Immunol. 39:995-1001(1995).

CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL RNA REPLICATION.

CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA CAPPING AND AN RNA HELICASE.

CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS BETWEEN CODONS FOR LEU-1112 AND GLN-1114.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Query Match 3.84; Score 339.5; DB 1; Length 1612;

Best Local Similarity 19.24; Pred. No. 1.2e-13;

Matches 351; Conservative 232; Mismatches 659; Indels 585; Gaps 85;

```

OY 94 LHRCE-----LRPYGRDYQWYSAPTRGPANCRSRLRGLP-----PADRTYC 137
Db 123 VHCMPENLDIRDAHRIHNOODTVST--LAKLERSK--RGLPVFQOQAFKYNMNDPVAVCC 179
OY 138 PDGFSRCAFAET-----GVALYSLHDLMPAD--VAEMARHGCTRLALHLPREVLTP 191
Db 180 DKRFQOCSYVDLPCKTAVYALHSTYDI--PADEFGAALLRKVHCICYAFHSENLL--- 236
OY 192 GYVHTTSTYL-----IHGDRAVVTYEGDTSAGYNHDVSLIRAMIRTT-----KIVG 238
Db 237 ---ETTSAPLDEIGATFYKSGDRLSFEFONESTLINEHSYKVKYKCTFEPASNRFYV 293
OY 239 DHPVIERARAGCHVLLT-----AAPEPSMRYV----- 271
Db 294 HKEFMCTRVNWFCKTKVDYTFLEFRGVYTRGDSQOFTAMDEAMEYKKTILAMKCEPT 353
OY 272 -----YPRSTEVVRSIF--GPGSPSLFSPACSTKSTF--HAVVHIWDRMLFG 318
Db 354 IFRDRAAVNFMFPKYKDMVIVPLFDSVYISGKMKRSEVNVNDVYTVLNIH----- 405
OY 319 AFLDOAFCCSLMTYLRGISYKVTYGVALVANEGMNASSEDALTAXITTAAYLTTCQRYLR 378
Db 406 RYQOKKALTYKKNVLSFEVESIRSRVILINGVASRSEMDVDSKYLQA--LSMTFL-----LQ 457
OY 379 TQALSGMRRLGVEHAQK-----ITRLYSWLFKESGRDIYFGROLQFYAOCRRMLSGFHL 435
Db 458 TT--LBAKQVYVKKFKOKFDPTVTNLEWKOISDAVGDLPESTK-----ETLISGGF-- 507
OY 436 DRYVLVDESVPCRCRTFLKRYAGKFCCEPMRLGOECTCFLEBAELVGDHGHNEAYEG 495
Db 508 ---VVAEGSLD-----KTRPEYITF-----ADKLWME-----YKA 536
OY 496 SEVDPAEPAHLDVSGTYAVHGHOLEALYNPODIAARASRLITAVELVASPDLECR 555
Db 537 TE---ELQHLDISKPL-----ERAEKYNAL-----SELSTYKE--COEFDITQFK 577
OY 556 TYLGNKTFRTYV-----DGAHLEANGPEQVYVLSFDASQSGAGSHSLTYELLTPA 606
Db 578 NICEEKDIDPPVAKYIYIMKNELTLPKNPPTPEALS--DA-----LSPDKR 623
OY 607 GLQVRISNGLDCTATFP-----PGAPSAAPGEVAAFCALRYRNFRTORSHSLTGILM 660
Db 624 DLDMRDLKLSTCAFPFVKTLDSGLPKQSYGDERQF-----ESQSVVSVSD 672
OY 661 LRPBGLIGTFPPSPCHIMESANPFCGEGTLYTRTWSTGFSFSDRPPAAPAMATPG 720
Db 673 FHLKSY-----ESYKMKSMSAVYTGPLKVOOMKNYM--DYLSASISAT-- 714
OY 721 LPHSTPPVSDIWLPPPESEFOVDAPVPAPDPAGLPGPVVLTTPPPPVHKSIPSPS 780
Db 715 VSNLCKVLADVYGADEPSAE----- 734
OY 781 RNRRLTYTPDGAKYVAGSLFESDCDMLVNASNPGHRRPGGGLCHAFYORFEAFYPTERT 840
Db 735 --KSGYDVVKG-----KWLKPK-----GKCHAM----- 757
OY 841 MREGLAAYLTTPRPIIHANAAPDYRVEQNKRLBAAYRETCSRRTGATAYPLLSSGIYQVPY 900
Db 758 ---GVALNNGEKVIV-----LELMA-----DGFICGD----- 783
OY 901 SLSDMERNHRRPDELTYLTF--PAANMFENKPNQPVLTITTEDTARTANLALDEIDAAHE 958
Db 784 -----WRVVAVSSDLSLYSDMKLOTLISLCKDGEPV----- 815
OY 959 VGRACAGCTISPGIYHOFYTAGVPGSGKRSIQG---DVDVYVVPVTRFLNSWRARG-- 1013
Db 816 -----PSDAKVTLVGVPGCGKTKELLETYNPEDLIVYGRKACAKHITIRAKK 864
OY 1014 -----FAFPTPTAARTYIGRRVVIDEAPSLPHELLLH-----MORA 1051
Db 865 SGHVRATKDNVFTVDFSLMHLKPKTY--NKLFIDEG-----LMLHTGCNVFLIALSHC 915

```

```

OY 1052 SSVHLLGDPNOIPAD-----FENAGLVPAIRBELAPTSMWXYTHRCPAD 1096
Db 916 REAMVEGDDEQLEFIRRVANFEPYPKHFAATLVYDHR---EVR-----RSLRCPAD 962
OY 1097 VCELLIGAAP--KIOTSRVLRSL-----FWN--BPATIGOKLVXTOAKA----- 1137
Db 963 VYHFMKSKDQKVLCTGNDVYIRSDAVYGVGKGFENKSKPDLKIKITTFQSDAKELKNG 1022
OY 1138 ---ANPGAI--TVHEAGATFETTTII--ATDARGLIQSSRAHAIYALTRHTEK---CV 1187
Db 1023 YEEVSFGEELNIVHEIOTGEFFEDVSIVRLTPPLELISKSSPVALVALTRHTEKSFYKVS 1082
OY 1188 ILDAPGLFREVG---ISDVYNNFELAGGEVXHHPSYIPRGNPDONIGTLQAFPPSQ 1243
Db 1083 VLDP--LVKVCSDLSKVSVDLIDMIVKVDAGLQLOVGSIFKG---ENL-----FVP--CP 1131
OY 1244 ISAYHOLAEELGHPAIVAAVLPCEPL--EQGLITM---POLTVSDSLVLELDIYH 1298
Db 1132 KSGY-----ISDMQFYUPTLLPGNSTILNEYDAVYMNLRNNLNVKDCIDFSKS---- 1181
OY 1299 CRMAAPSOBKAVLSTLVGRGRTKLYEAHSDVRESLARFIPT--IGPVQATTCELYEL 1356
Db 1182 --VSVPRQOOEFTFPVY-----RTAAERPRSRGILENLVAMIKRNFNSDILTDIEDPT 1234
OY 1357 YEAMVEKGDDGSAVLELDCNRPVSRITFEQKXC---NKFTGETIANGKVG----- 1405
Db 1235 AELVNVKFWDAYITIDLSGQNTVPMTSDAFHMMAKOEKSTIGQLADDFVDLPAIDQYK 1294
OY 1406 -----OGISAWSKTEFCALGPMFRAIEKILALLPNN--IFYG 1441
Db 1295 HMIKAQPKQKLDLSDPODEYAAQOTIYHKSQINAIIGPLSELTROLLERLSDSKFLFT 1354
OY 1442 DAYEESV--FAAIVSAGSACOMFENDSEFSDTONNPSGLGECVWHEGMPQWLIRLYH 1499
Db 1355 RKTPEQIEEFSDLDSTVMEVYELDISYKDKQNEFHCAVEYLIMEKLTGNGFL----- 1409
OY 1500 LVRSAMILQAPKESLKG-----WKHSGEPGTLMTNTVMNMAIIAHCEFEFDPRY 1550
Db 1410 ---EYVKQGHKRTSLMDYTAGITCMTQKSGDVTTFIGNYITIAACLSMTPMDKVK 1467
OY 1551 AAFKGGDSVYVLS--DYROSNAALIAAGCGILKLVDRP--IGLYAG--VVAAPGLGTLTP 1605
Db 1468 AAFCGDSDMLYIKGLDLPDIOGANLMMNFEAKL--YKRRGYFCGRRIIHNDGALIV 1524
OY 1606 --DVYAFAGRLSKMNGPPEBAEQRLAVCDPLRGITLWNAOV--CYDVYSKRYGVSPGL 1661
Db 1525 YYDPVKLISLGCXKI--KSLDHLKEFRISLSDVSASLNCAYYGOLDATAIEVH----- 1577
OY 1662 VHNLIQMLOTIADGKAHFTETIKPVL 1688
Db 1578 -----KITAVNGSFAFCSYVKYLS 1596

RESULT 11
RPO.TOML STANDARD: PRT: 1616 AA.
AC P03587; 041352;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) (CONTAINS:
DE METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)).
OS Tomato mosaic virus (strain L) (TOMV) (TMV strain tomato).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=12252;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85157522; PubMed=6549393;
RA Ohno T., Aoyagi M., Yamashita Y., Saito H., Ikawa S., Meshi T.,
RA Okada Y.;
RT "Nucleotide sequence of the tobacco mosaic virus (tomato strain)
RT genome and comparison with the common strain genome.";
RL J. Biochem. 96:1915-1923(1984).

```

CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPING AND AN RNA HELICASE.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X02144: CAA26085.1; -
DR EMBL: X02144: CAA26082.1; -
DR PIR: A04195: WMTMT.
DR Interpro: IPR000606; -
DR Interpro: IPR001788; -
DR Interpro: IPR002588; -
DR Pfam: PF00978: RNA_dep_RNAPol2; 1.
DR Pfam: PF01443: Viral_helicase1; 1.
DR Pfam: PF01660: Ymethyltransf; 1.
DR Transferrase: RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
FT NE_BIND 833 840 ATP (POTENTIAL).
FT SEQUENCE 1616 AA; 183564 MW; A8BC8929B5CFCAC CAC64;

Query Match 3.68; Score 322.5; DB 1; Length 1616;
Best Local Similarity 19.2%; Pred. No. 1.5e-12;
Matches 349; Conservative 205; Mismatches 660; Indels 601; Gaps 79;

QY 91 PNVLHRCFLRPGVGR--DVGRWYSAFTRG--PAANCRSAL-RGLPRADRTYCFDGFSCA 145
DB 128 PNLVDVRIMHREGOKDIEYLSRLERGNKHVPFOKEADRYAEMNEVYCHDTFOTCR 187
QY 146 FAAE--TG-----VALYSLHDLMPAD-VAEMARHGXTRLVYALILPREVLLPG--TYHT 196
DB 188 HSQECYGRVYALALHSITDI-PADEFGALLRKNVCHVCAAFHFSNLLLEDHVNIDE 246
QY 197 TSYLLIHGDRAVVTYGDTSAGYNHDVSLIRANIRTKTKIVGDHPLVIERVRAIGCHFVL 256
DB 247 INACFOGDRLETFESFSESTLANSYSNLIKVKYCKT----- 284
QY 257 LITAAPRSPMPYVPRSTREYVYRSIFGPGSPSLFPPSACSTSTPHAVPHIMDKML 316
DB 285 -----YFP-ASNREVIYKE-----FLVTRVNTW----- 306
QY 317 FGATLDDQAF-C SRLMTYL--RGISYKVTYGVALANEGMNASEDALTAITTAAYLTICH 373
DB 307 -----FCKSRIDTFILLYKGVAKH-----GVDSQDFYKAMDAMHYKKT--LAMCN 350
QY 374 -QRLKTOAISKG-----MRRLGYEHAOKFTRLYSWLFKSGR-----DYIFGRQ 418
DB 351 SEBILLEDSSSVNWPFKMDM-----VIVPLFDISLETSKTRKREVLVSKDFV----- 399
QY 419 LQFYAOCR--RWLSAGPHLDPRVLVEPVSRCRTFLKKYVAGFCGFMRLGECFCFL 476
DB 400 -----YTULNHRTTQAKKALTYSNVLSVEST--NSRVIIINGVTAR----- 438
QY 477 EPAEGLVGDHNDNEAGESEVDPAPPAHLDVSGTYAVHGHOLEALYRALNVPDIDARA 536
DB 439 -----SEMDVDKSLQSLSMTEFLH----- 458
QY 537 SRLTAIVELVAPDRLECRVVLGNKTRFTTVVDGAH----- 572
DB 459 -----TKLAVLKDDLLISKRFALGPKTYSOHWMEISLAFGAPPSIKERLINRKLKITE 513
QY 573 --LEANGPEGYVLSFD--ASROSGAGASHL-----TYELTPA--GLQVRISSGGLD 618
DB 573 --LEANGPEGYVLSFD--ASROSGAGASHL-----TYELTPA--GLQVRISSGGLD 618

DB 514 NALRIYVDDLYVTPHDLVSEYKMSVDMPLYDIRKKMEETEEMYNALSELVLKNSDKFD 573
QY 619 CTATPPPGAGAPSAAPGVAAFCSALYRNRTFOGHSLLTGGIMLHPBELGLIFPPSPGHI 678
DB 574 VDV-----FSQMCOSL----- 584
QY 679 WESANPCGEGTLYTRTWSTSGFSSDSFPEAAPAMAAAPGLPHSTPPVSD--IWVLP 736
DB 585 -EVDPMIAKVIYAVVNSGSLTLEFQPTAEANVALAL-----QDEKMSDGLAVYTSR 637
QY 737 PSEFQVDAAPVPPAPPAAGLPGPVYLTTPPPPVHKPSIPPSRNRLLYTPDCAKY 796
DB 638 DVEPSSIKSGMARGELQALGSDV-----PES-----SYRSEIE 674
QY 797 AGSLFEESDCMLVANSNPGHPPGGGLCHAFYQRPFAFPTEFIMRGLAAYTLTPPPI 856
DB 675 SLEQFH-----MATASSLIHQ--KCSIYTYGKLKQOKKNFI--DSLVAASAAVSNL 724
QY 857 HAVAPDYRVBQNPKRLEAAVRETCSRRTAAYPLLGSIYQVPVLSFDAM--ERNHRPG 914
DB 725 VKILKOTAAIDLETROKFGVLDVASKR-----WLVPKSAKNHAGVETHARK 772
QY 915 DELYTE-----PAANPEANKPAQPVLTITEDTARTANLAIEDAATEVGRACAGCT 967
DB 773 YHVALLEHDEFGIITCDNMRVAVSSSVYSDMAKLRTLRLLK-DGEPHVSSA----- 826
QY 968 ISPGIVHOFYAGVPGSGKRSISIQG--DVDVVVVPVTRELNSMRRR--GFAAFTPH 1020
DB 827 -----KVVLVDVPGGCGKTKKELSRVNFEDLILVGRQAAEMIRRRANASGIIVATKD 880
QY 1021 TAARV-----TIG-----RRVVIDEAPSLPHLLH-----MORASSVHLGD 1059
DB 881 NVRTVDSFLMNYGARGCOFRLEFIDG-----IMLHGCNVFLVEMSLCIDAYVYGD 933
QY 1060 PNOIPALIDFENAGLVPA--IRPELAPTSMMXVYHRCADVCGLRGAYP-KIOTSVLR 1116
DB 934 TQOLPIYINRYGFPYPAHFAKLEDEVETRRRTLRCPADVTYHFNQRIEGHVMCTSEKK 993
QY 1117 SLEW-----NEPAIGOKLYVXTOAKAA-----NPGATITHEAGATFTETII 1159
DB 994 SVSGEMVGAASINPVSKPLKGLITFTQSDKEALLSRGADVITVHEVGETYADVSLV 1093
QY 1160 -ATDARGLIOSSRAHAIVALTRHEK-----CYLDA-----P 1192
DB 1054 RLPTPTVSIIAROSPVALVLSRHTKSLKYTVVMDPLVSIIRDLERVSSYLDMYKVA 1113
QY 1193 GLIREVGISDIYV-NPEFLAGGEVGHHRPSVYIPRGNDOMLGTLOAPPPGOISAHYOLA 1251
DB 1114 GTOXQLOVDVSFFKNFNLVPAAPKTG-----DISDQFYDKC----- 1150
QY 1252 EELGHRPAPVAAVLPCEPELEOGLIYMPOLIVSDS--VLVFEITDI-----VHC 1299
DB 1151 -----LPGNSTLLNNVDATYMKLTIDISLWVKCITLMSK 1184
QY 1300 RMAAPSGRAVLSTLVGRYGRFKLYEAHSDVRESLARIPY--ICPVQATTCELYELV 1357
DB 1185 SVAAPKDVKPTLLPMV-----RTAAEMPROTGLENLVAMIKRNFNSPELSGVVDIENTA 1239
QY 1358 EAMVEGODSAYVELDLNCRNDVSRITFPQKXCKF--TTEGTIHAGVY----- 1405
DB 1240 SLVYDKDFDSYLKEKKPKPKNS--LFSRESLNMTAKQEOYVITGLADPFDVLDPAVD 1297
QY 1406 -----OGISAMSKTFCALFGPWRARIEKITALL----- 1434
DB 1298 OYRMIKAOPKOKLDSIQTEYPALQTIYVHKKINAIFFPLPSLRLDIDSISRFL 1357
QY 1435 -----PPNI--FYGDAYEESVRAAVSAGSGSCWYFENDSESDSTONNLSLGEVYME 1486
DB 1358 FTRKTPPAQIIEDEFGD-----LDSHVPMQVLELDVSKYDKSNEFHCAVEYEIWR 1407
QY 1487 ECGPQWLIRLHYVRSAMWLOAPKESLKGf-----MKHSGSEGPLMMNTVNMMA 1537
DB 1408 RLGEDPFLAEV-----WQGHKRTTKLDKYTAGIKTCLMAYORKSGDVTTFGTAVIIAS 1460

QY 1538 IIAHCEPDEFERNAKGDSDVVL-----CSDYRQSRNAALAGCGLK-LKYDYRPIGLY 1592
DB 1461 CLASMLPMKELKIGAFCCGDDSLLYPEKGC-EXPDIQQAAMNMFKAJFKKQY---GYF 1516
QY 1593 AGVYVA---PGTGTDPDVVAFAGRLSE---KNMGCPGEARQOLRLAVCDLRLGTVAOV 1646
DB 1517 CGRYVHHHRCGLVYDPLKLSKLGAKHKKDM---DLLEFRRLCDVAESLNACAY 1572
QY 1647 CV--DVYSRYGVSP 1659
DB 1573 TQLDVAQGEVHTAP 1587
RESULT 12
RRPO_TOMSI STANDARD: PRT: 1616 AA.
ID RRPO_TOMSI
AC Q9YKD6: Q9YJ37;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE RNA-DIRECTED RNA POLYMERASE (EC 2.7.7.48) (183 KDA PROTEIN) [CONTAINS:
METHYLTRANSFERASE/RNA HELICASE (MT/HEL) (126 KDA PROTEIN)].
OS Tomato mosaic virus (strain S-1) (TOMV).
OC Viruses: ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_Taxid=138314;
RN (1)
RP SEQUENCE FROM N.A.
RA Zhou X., Xue C., Chen Q., Qi Y., Li D.;
RT "Complete nucleotide sequence of a Chinese isolate of tomato mosaic
RT virus";
RL Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPING AND AN RNA HELICASE.
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: AJ132845: CAB36997.1; -;
DR EMBL: AJ132845: CAB36998.1; -;
DR InterPro: IPR000606; -;
DR InterPro: IPR001788; -;
DR InterPro: IPR002588; -;
DR Pfam: PF00978: RNA_dep_RNApol2; 1.
DR Pfam: PF01443: Viral_helicase1; 1.
DR Pfam: PF01660: Vmethyltransf; 1.
DR Transferrase: RNA-directed RNA polymerase; Helicase; ATP-binding.
KW CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
FT NP_BIND 833 840 ATP (POTENTIAL).
SQ SEQUENCE 1616 AA; 183542 MW; 5DBBFB2FADCC5C0C CRC64;

Query Match 3.6%; Score 321.5; DB 1; Length 1616;
Best Local Similarity 19.4%; Pred. No. 1.7e-12;
Matches 352; Conservative 218; Mismatches 649; Indels 591; Gaps 83;

QY 91 PNLVHCFIAPVGR--DVQRWYSAPTRG--PANCRRSL-RGLPRADRYTCGDSRCA 145
DB 128 PNLVDVDMRHEGQKDSIEVLSRLERKNKHNVEFQKAEFDRAEPNEVYCHDTQTCC 187
QY 146 FAEE--TG-----VALYSLHDLMPAD--VAEMARHGXTRLYLALHLPPEVLLPPG--TYHT 196
DB 188 HSGECVTGVAVALHLSYDI-PADEGAALLKKNVHVCYAAHFPSBNLLDESHVND 246

QY 197 TSYLLIHDDBRAVYVEGDTSGAGYHNDVSLRAMITTKIVGDHPLVIERVRAIGCHEVL 256
DB 247 INACFORDDRLTFPSASESTLNYSHSYSNILKYCKT----- 284
QY 257 LITAAPEPMPVPVPRSTEVYVRSIFPGGSPSLFSPASCTKSTFHAHPVHMDRLM 316
DB 285 -----YFP-ASRREYIMKE-----FLYTRVNTW----- 306
QY 317 FGATLDDQAFCC-CSRLMTYL--RGISKYVTGALVANEGMNASEDALTXATAVLTIGH 373
DB 307 -----FCKFSRIDFTLYKGVANH-----GVDEQYTKAMEDAMHYKKT---LAMCN 350
QY 374 -QRYLRTQALSKG-----MRRLGVEHAQFTRILYMLEKSGR-----DYLPGHQ 418
DB 351 SERILLDESSVYVWPCKRDM-----VYPLFDISLETSKRTREVLVLKDFV----- 399
QY 419 LQFYACR--RMLSAGFHLDPRLVDFDESVPCCRFELKVKVAGKFCFCFMRMLGQECTFL 476
DB 400 ---YTLNHIPTTYQAKALYTSNLSFEVSI--RSRYIINGVTAR----- 438
QY 477 EPAEGLVGDHGDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNPQDIARA 536
DB 439 -----SEMDVDRSLQSLSMTEFFL----- 458
QY 537 SRLTATVEL-VASPDRLCEKTYVGNKTFRTYVDGAH----- 572
DB 459 -----IKLAVLKDDLILSKFALPKPTVSOHWDEISLAFGNAFPIKERLINRKLKIT 512
QY 573 ---LEANGPEQVYLSFD--ASRQSMGAGSHL-----TYELTPA--GLQVRISNG 617
DB 513 ENALETRVDPVLYTFHDLRYSEYKMSVDMPYLDIRKMEETEMVYALSETLVKKSDF 572
QY 618 DCTATPPPGAPSAABGEVAFAFCALYRYNRFTQRSLTGLVHLEGGLGTFPPSPGH 677
DB 573 DVDV-----FSQMQLS----- 584
QY 678 IMESANPFCGEGTLTYRTMSTSGFSSDFSPPEAAAMAATPCLPSTPPVSD--TWLP 735
DB 585 ---EVPMTAAKVIYVMSNESGLTLTFPDPTFANVALL-----ODSEKASGALVVS 636
QY 736 PSEEPQVDAAPVPAPDPDAPGVPVLPVPPRPVHKSIIPPSNRRLYPPGAVY 795
DB 637 RDVEEESIKSGNARGLQLAGLSGV-----PES-----SYTSEEL 673
QY 796 YAGSLFESDCDMLVNASNPGHPRGGGLCHAFYORFEAPVPTFT----- 840
DB 674 ESLEQFH-----MATASSLIHQ-----MCSIVYIGPLKYQCMKRFIDSLVASSAANSNV 725
QY 841 -MREGIAAYTLPRP---IIHAVAPDYRVEQNDKRLAAVRETCSRGTAAVPLGSGTY 896
DB 726 KILKDTAAIDLETROKFGVLDVASKRMVLKPSAKNAMGCVETRHAKYVA--LLEHDEF 783
QY 897 QVYVSLFPAWENHNRPGDELILTEPRAANFEAKRAQVLTITTEPTARANT--ALEIDA 955
DB 784 GI---ITCDWRR-----VAVSPESVYSOMAKIRTLRLKLG 819
QY 956 ATEVGRACAGCTISPGIVHGYFTAGVPGSGKRSIQG---DVDVAVVPTRELNSWR 1012
DB 820 EHVYSSA-----KVLVLDGVPGCGKTKELLSRVNFEEDLLILVPGQAAMIR 868
QY 1013 ---GFAFTPTTAAV-----TIG-----RRVVIDEAPSLPULLLH----- 1047
DB 869 ANASGIIVATKDVVRVDSFLMNYGARGARQCFRFLIDE-----LMILHTGCVNFVE 921
QY 1048 MRASSVHLIGDPNOIPALDFEHAQLVPA--IPELAPTSWMVYTRCPADVCCELLRGAY 1105
DB 922 MSLSDIAVYIGDTQQLPYINRVTGPPYPAHFALVEDEVETKRTTLRCPADVTHFINOR 981
QY 1106 P-KIQTTSRYLNSLFW-----NEPAIGQLVYVTOAKAA-----NEGATVHE 1147
DB 982 EGHVMTSSSKRSVQSEMVSGASINPVSKPLGKILTLTFQSDKALLSGYDVHTVHE 1041

Db 782 ELPCDADMSKVAVSNESMV----- 801
QY 784 KLLIYPPGAKYVAGSLFESDCDMLVNASNPGHRGGGLCHAFYORFPFAFPTFERIME 843
Db 802 -----YSDMAKL----- 808
QY 844 GLAATVFPRIIHAVDPYVEQNPKRLAAYRETCSRGTAAVPLIGSGIYOVPVSL 903
Db 809 -----RVLKRSIGEMPIYSV 823
QY 904 FDAMERNHRPGDELTYLTPANMFANKPAOPVLTTEDTARTANLALEIDAATEVGRAC 963
Db 824 -----SAKYTL-VD----- 831
QY 964 AGCTTSPGIVHQTAGVPGSGKRSIQG---DVDVYVVPTRRLNSWRRRGFAFTPH 1020
Db 832 -----GVPCCGTREILRRVNSEDLVLPGEKAAMLRK--ANOSGN 873
QY 1021 TAAR-----VTG-----RRVVIDEAPSLPPLL--LTHMQRASSVHLGDP 1060
Db 874 IVANNONKTYDSFLMNLGKPGVCOFKLEVDEGLMHPGCVFLKLSLCEAFVFGDT 933
QY 1061 NOIPAIDE-----EHAGLVPARPELAPTSMWXYTHRCPADVCELIRGAYP-KIQTTS 1112
Db 934 QOLPIFNRYQNFPPOHFSKLVDEFEKRT-----TLRCPDVTHFLNQCYDGAVTTS 988
QY 1113 RVLBSLFW-----NEPAIGOKLVYTOAK-----AANPGATVHEAGATFTE 1155
Db 989 KTORSGLEVGVGAAMNPFVKPLGKIYTFQSDKLTMLSRGODVNTVHEIOGETYEE 1048
QY 1156 TTII-ATDARGLIQSSRAHAIVALTRHEKC-----VILD-APGLREVGISDVIYVNF 1208
Db 1049 VSLVRLTPPTTHIISRESHPVGLRHT-KCFKYTYVLDPLVKLVRLD---ECVSNF 1103
QY 1209 FL-----AGEVGAHRPSYIPRGNPDNLGTLOAFPPSCQISAYHOLAE 1253
Db 1104 LLDVYMWDSVSAKXOVSGVLAENLFVQAPKSGDQD---LQFYDKCLPGNSTVLNE- 1159
QY 1254 LGHRAPAVAVLPCEBELOGLLYMBOELTVSDSVLFE-----LTDIVHCR 1300
Db 1160 -----FDAVTMNCSDI-----SLNVKDCVLDFSKSVPLPRDNTKVPMTPIV--R 1201
QY 1301 MAPSOR--KAVSTLVG-----RYGRRTKLYEAH 1329
Db 1202 TAAERPRSGGLEENLYAMIKRNFNSPELSGYDMENTASVADRFDSYFLKDKLSGCSL 1261
QY 1330 SDV-----RESLARFI-----PTIGPVQATTCELYELVE-----AMVEKGODGSAY 1370
Db 1262 GDSGCKNIIDRQALLRMMEKQEKSTIGLAD-----YDFVDLPAIDQYRHIIKSOBKQ-- 1314
QY 1371 LELDLGNRDVSKITTFQKCKNFTTGETTAHGKVGOGISAMSKTCALFGPMFRAIEKEI 1430
Db 1315 -KLDLSIQ-----SEYPSLOFTIVY-----HSKKINALPGPIFSELTROM 1352
QY 1431 LALLPN--IFYGDAVEESV--FAAAYSAGSOMFENDESEFDSQNNFSLGLECVME 1486
Db 1353 LSAIDTSKLTFTKRPEDIEEFESDLDAHQPMVEVLELDVSKDQSONEHCAYETEYWK 1412
QY 1487 ECGMPQMLIRLYHLVNSAMIIQAPKESLKG-----WKHSGEPGLTLMNTVWMA 1537
Db 1413 RLGIDEFLEAV-----WKQGHRTKTLKDYTAGIKTCLMYQOKSGDVTTFIGNVTIAA 1465
QY 1538 ITAHCEYFDFRVAAFKGDSDVYLCSDIQSRNAALLIAGCGL-----KLKV 1584
Db 1466 CMASMLPMEKVIKAACFGDSDLVY-----LPKGCCELPTQSCANIMMNFBAKL 1513
QY 1585 DYRPIGLYAGVVVA---PELGLTLPDVVPAPAGRLSEKNMGPGPERAQLRLAODFLRGLT 1641
Db 1514 FKRTGYFCGRYVIHHDRAIYVYDPLKIISKIGAKHT-TDKEHLEEFRIISLADVSKSLN 1572
QY 1642 NVAOYCV--DVVSRYVYVSP 1659
Db 1573 NCAYYAQLDEAVREVHKTA 1592

Search completed: May 30, 2001, 16:14:36
Job time: 288 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 16:06:58 ; Search time 43.81 Seconds

(without alignments)
2663.582 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013

Sequence: 1 PGTTTAEQALAAANSALA.....FTETIKPVLDLINSIORVE 1698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR1:*
2: PIR2:*
3: PIR3:*
4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7448	82.6	1693	1 MNMWH	genome polypeptide
2	7433	82.5	1691	1 A44212	genome polypeptide
3	1368	15.2	290	2 C48547	nonstructural prot
4	763	8.5	152	2 B38196	probable RNA-dirc
5	391.5	4.3	1646	1 MNMWS2	186k protein - cuc
6	390.5	4.3	2115	2 S38480	nonstructural prot
7	374.5	4.2	2205	1 MNMVRN	nonstructural poly
8	372	4.1	141	2 A38196	non-structural hyp
9	349	3.9	1844	2 S01956	hypothetical prote
10	346	3.8	1985	2 S19151	hypothetical prote
11	324	3.6	1616	1 J02144	180k protein - tom
12	323	3.6	1615	1 MNM8T	183k protein - tom
13	320	3.6	1597	2 S65053	genome polypeptide
14	314	3.5	1601	2 S48699	178k protein - tob
15	313	3.5	1611	1 MNMTPV	genome polypeptide
16	311	3.5	1748	1 J01555	genome polypeptide
17	310	3.4	1844	1 RRMPTM	genome polypeptide
18	305.5	3.4	1776	1 RRMPTM	genome polypeptide
19	296	3.3	1615	1 MNM18	183.3k protein - t
20	293.5	3.3	1839	1 RRMPEM	genome polypeptide
21	293	3.3	1874	1 J00533	genome polypeptide
22	260	2.9	1608	1 MNMGM	183k protein - tob
23	243.5	2.7	1707	2 S01865	genome polypeptide
24	243.5	2.7	1884	1 A45353	genome polypeptide
25	242.5	2.7	1885	2 J02183	hypothetical 216.5
26	227	2.5	1112	2 S49432	replicase 126k - o
27	226.5	2.5	1766	2 S03701	141k protein - pea
28	218	2.4	2512	1 MNMVS	nonstructural poly
29	211	2.3	2514	1 MNMVB2	nonstructural poly

30	207	2.3	2514	1 MNMVB2	nonstructural poly
31	202.5	2.2	2492	1 C44213	nonstructural poly
32	202	2.2	1729	2 A49282	fusion protein 1a/
33	201.5	2.2	1707	2 S77910	hypothetical prote
34	201.5	2.2	1718	1 J01734	genome polypeptide
35	200	2.2	2431	1 MNMVSF	nonstructural poly
36	194	2.2	1968	1 PNO093	genome polypeptide
37	184.5	2.0	3175	1 RRMVEV	hypothetical prote
38	184	2.0	1707	2 S77908	hypothetical prote
39	183	2.0	1116	2 S26358	nonstructural poly
40	177.5	2.0	2492	1 MNMVTD	nonstructural poly
41	176	2.0	1385	2 S34230	156k protein - pia
42	174.5	1.9	2492	1 A44213	nonstructural poly
43	172.5	1.9	2493	2 S26372	nonstructural poly
44	169.5	1.9	2493	2 S72349	nonstructural poly
45	167	1.9	810	1 P2WMBB	2a protein - broad

ALIGNMENTS

RESULT 1

MNMWH

genome polypeptide - hepatitis E virus (strain Burma)
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.48)

C:Species: hepatitis E virus
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 23-Jul-1999

C:Accession: A40778; A48547
R:Tam, A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes, Virology 185, 120-131, 1991

A:Title: Hepatitis E virus (HEV): molecular cloning and sequencing of the full-length
A:Reference number: A40778; MUID:92024067

A:Accession: A40778
A:Residues: 1-1693 <TRAM>

A:Cross-references: GB:M73218; NID:9330023; PIDN:AAA45734.1; PID:9330024

A:Note: sequence extracted from NCBI backbone (NCBIN:104572, NCBI:P104573)
A:Title: Hepatitis E virus (HEV): strain variation in the nonstructural gene region e

A:Reference number: A48547; MUID:92271462
A:Accession: A48547

A:Molecule type: genomic RNA
A:Residues: 967-1693 <ERY>

A:Cross-references: GB:M32400; NID:9330021; PIDN:AAA03206.1; PID:9330022
A:Note: sequence extracted from NCBI backbone (NCBIN:104572, NCBI:P104573)

C:Superfamily: hepatitis E virus nonstructural protein
C:Keywords: ATP; nonstructural protein; nucleotidyltransferase

Query Match 82.6%; Score 7448; DB 1; Length 1693;
Best Local Similarity 81.8%; Pred. No. 0;
Matches 1396; Conservative 110; Mismatches 169; Indels 32; Gaps 6;

QY	1	PGTTTAEQALAAANSALANAVVPRPLSRVOTELLINMOPROVPEVIMNPIOR	60
DB	10	PGTTTAEQALAAANSALANAVVPRPLSHOIELINLMOPROVPEVIMNPIOR	69
QY	61	VHNELEOYCRARAGRCLEVGAFHSINDPNVLRHCFLPVGRDQVRYSAFTRGPAN	120
DB	70	VHNELEOYCRARAGRCLEVGAFHSINDPNVLRHCFLPVGRDQVRYSAFTRGPAN	129
QY	121	CRSARLRGIPADRTYCFGFSKCAFAETGVALYSLHDMRPDAVEAMARHXTLYAA	180
DB	130	CRSARLRGIPADRTYCFGFSKCAFAETGVALYSLHDMRPDAVEAMARHXTLYAA	189
QY	181	LHLPPEVLLPFGYHTSYLLIHGDRAVYTYEGDSAGYNHDSILTRAMIRTKTVGDH	240
DB	190	LHLPPEVLLPFGYHTSYLLIHGDRAVYTYEGDSAGYNHDSILTRAMIRTKTVGDH	249
QY	241	PLVIEFVRAIGCHFVLLTAAPEPSPMPVYPRSTEVYVRSIFGSGSPFLPSPACSK	300
DB	250	PLVIEFVRAIGCHFVLLTAAPEPSPMPVYPRSTEVYVRSIFGSGSPFLPSPACSK	309

```

QY 301 STEHAIVHIDRLMFCATLDDOAFCCSRMLTYLRGISYKVYGVAGLVANEGMNASDAL 360
D 310 STEHAIHVIDRLMFCATLDDOAFCCSRMLTYLRGISYKVYGVAGLVANEGMNASDAL 369
QY 361 TAXITAAVLTICHOXYLTOAISKMRRLGVEHAKETITRYSWLFEEKSGDYTPGRLO 420
D 370 TAXITAAVLTICHOXYLTOAISKMRRLGVEHAKETITRYSWLFEEKSGDYTPGRLO 429
QY 421 FYAOCRRMLSGFHLDPRLVFEDESVPCCRTFLKRVAGKRCCEMRWMOGECTCLPEAE 480
D 430 FYAOCRRMLSGFHLDPRLVFEDESAPCHCRALRKALSKCCCFMKWMOGECTCLPEAE 489
QY 481 GLVGDHGNDAVBSSEVDPAEPALNDVSGTYAHQHOLEALYALANVPODIARASRLT 540
D 490 GAVGOGGNDNAVBSSEVDPAEPALNDVSGTYAHQHOLEALYALANVPODIARASRLT 549
QY 541 ATVELVASPDLECEVTYGNKTFRTYVDGAHLEANGPEOYVLSNDASROSGSGSHLT 600
D 550 ATVELVASPDLECEVTYGNKTFRTYVDGAHLEANGPEOYVLSNDASROSGSGSHLT 609
QY 601 YELTPAGLOVRISNGLDCTATFPFGAPASAPGEVAFAFCALYRNFTORHSLTGLM 660
D 610 YASASAGLEVYVAAGDHRARAVFAPGVSRSAPGEVATFCALYRNFTORHSLTGLM 669
QY 661 LHPGLLGIFPPSPBGHIESANPCGEGTLYRTWS-TSGFSSDFSP----PEAARA 714
D 670 FHPGLLGIFPPSPBGHIESANPCGEGTLYRTWS-EVDAPEADLGFMSPEPIS 729
QY 715 MAATPGLPHSTPRVSDIMVLPPEPSEEPYDAAV-PPADPAAGLP-GVVLTPPPPPVH 772
D 730 RAATPPLT-----AALPPPADSPSPSAPALALPPASGATAG 766
QY 773 KPSP-PPSRNRRLTYTPDCAKAYAGSLFESDCLVNAANPCGHPGGLCHAFYORFP 831
D 767 APALTTHOARHRLLETPDGSKVYAGSLFESTCWLVAANVDRHRCGGLCHAFYORFP 826
QY 832 EAFPTTEIMEGLAATYLPRLPIIHAVAPDYRVEONPKLEAAYRETCSRGTANPPL 891
D 827 ASFPAASVVMDDGAAYLTPRLPIIHAVAPDYRVEONPKLEAAYRETCSRGTANPPL 886
QY 892 GSGIYOVSLSFPAWENHRPDELTYLTPAANWFEANKRPOVLTTEPTAPRANAL 951
D 887 GTGIYOVSLSFPAWENHRPDELTYLTPAANWFEANKRPOVLTTEPTAPRANAL 946
QY 952 EIDATEVGRACAGCTISPGIVHOFTAGVPSGSKRSIOGDVYVVVPTRELNSMR 1011
D 947 EIDATEVGRACAGCTISPGIVHOFTAGVPSGSKRSIOGDVYVVVPTRELNSMR 1006
QY 1012 RGFPAFTPHTAARTYIGRRVYIDEAPSLPHLLHLMORASSVHLGDPNOIPAIDEHA 1071
D 1007 RGFPAFTPHTAARTYIGRRVYIDEAPSLPHLLHLMORASSVHLGDPNOIPAIDEHA 1066
QY 1072 GLVPAIRPELAPTSMXXXTNRCPADVCLIRGAYPKIOTTSRVLSRLFWNPAIGOKLVX 1131
D 1067 GLVPAIRPELAPTSMXXXTNRCPADVCLIRGAYPKIOTTSRVLSRLFWNPAIGOKLVX 1126
QY 1132 TOAKAANPGATVHEAOGATFETTTIATADANGLIQSSRAHAIVALTRTEKCVILDA 1191
D 1127 TOAKAANPGATVHEAOGATFETTTIATADANGLIQSSRAHAIVALTRTEKCVILDA 1186
QY 1192 PGLIREGISIVYNNFPLAGCEVGHRSYIPRGNPNONIGTQAAPPSCQISAHYOLA 1251
D 1187 PGLIREGISIVYNNFPLAGCEVGHRSYIPRGNPNONIGTQAAPPSCQISAHYOLA 1246
QY 1252 BELGHRAPVAAVLPPELEOGLLYMPQELTVSDVLYFLTQIVICRMAAPSORAVL 1311
D 1247 BELGHRAPVAAVLPPELEOGLLYMPQELTVSDVLYFLTQIVICRMAAPSORAVL 1306
QY 1312 STLVGARYGRATKLYEAASDVRESLARFIPTIGVQATTCLELYEAMVEKGGDSAVL 1371
D 1307 STLVGARYGRATKLYEAASDVRESLARFIPTIGVQATTCLELYEAMVEKGGDSAVL 1366
QY 1372 ELDCNNDVSKITTFQCKCKNFTTGETIAHGKVGOGISAMSKTFCALGPPWFAIEKEIL 1431

```

```

D 1367 ELDCNNDVSKITTFQCKCKNFTTGETIAHGKVGOGISAMSKTFCALGPPWFAIEKEIL 1426
QY 1432 ALBPNIIFYGDAVESVPAVAAVYVPEFNSPESPDSONNPFSGLECVVEECMP 1491
D 1427 ALBPNIIFYGDAVESVPAVAAVYVPEFNSPESPDSONNPFSGLECVVEECMP 1486
QY 1492 QMLIRLYHLVRSAMLIQAPKESLKGFMKHSGEPTLMTVMNAALIAHCEFRDFEFA 1551
D 1487 QMLIRLYHLVRSAMLIQAPKESLKGFMKHSGEPTLMTVMNAALIAHCEFRDFEFA 1546
QY 1552 AFKGDSDVYLCSDYROSNAALTAGGGLKVDPRPLGLYAGVAVAGLCTLPDYVFA 1611
D 1547 AFKGDSDVYLCSDYROSNAALTAGGGLKVDPRPLGLYAGVAVAGLCTLPDYVFA 1606
QY 1612 GRLEKMWGPEPERAEOLRLAVCPDLRGTLNVAOVVYVRYGVSPGLVHNLIGMLOT 1671
D 1607 GRLEKMWGPEPERAEOLRLAVCPDLRGTLNVAOVVYVRYGVSPGLVHNLIGMLOT 1666
QY 1672 IADKAHETETIKPPLDLTNSIIORVE 1698
D 1667 IADKAHETETIKPPLDLTNSIIORVE 1693

```

```

RESULT 2
A44212
genome polypeptide - hepatitis E virus (strain Mexico)
N:Contigs: RNA-directed RNA polymerase (EC 2.7.7.48)
C:Species: hepatitis E virus
C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 23-Jul-1999
C:Accession: A44212; B48547
R:Huang, C.C.; Nguyen, D.; Fernandez, J.; Yun, K.Y.; Fry, K.E.; Bradley, D.W.; Tam, A.
Virology 191, 550-558, 1992
A>Title: Molecular cloning and sequencing of the Mexico isolate of hepatitis E virus
A:Reference number: A44212; MUID:93079857
A:Accession: A44212
A:Molecule type: genomic RNA
A:Residues: 1-1691 <HUA>
A:Cross-references: GB:W74506; NID:q330017; PIDN:AAA45730.1; PID:q330018
R:Fry, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Platak, M.; F.
Virus Genes 6, 173-185, 1992
A>Title: Hepatitis E virus (HEV): strain variation in the nonstructural gene region e
A:Reference number: A48547; MUID:92271462
A:Accession: B48547
A:Molecule type: genomic RNA
A:Residues: 965-1691 <FRY>
A>Note: sequence extracted from NCBI backbone (NCBITN:104576, NCBIIP:104578)
C:Superfamily: hepatitis E virus nonstructural protein
C:Keywords: ATP; GTP binding; nonstructural protein; nucleotidyltransferase; P-loop
F:973-980/Region: nucleotide-binding motif A (P-loop)
F:979/Binding site: ATP/GTP (Lys) #status predicted

```

```

Query Match 82.58; Score 7433; DB 1; Length 1691;
Best Local Similarity 81.08; Pred. No. 0;
Matches 1384; Conservative 121; Mismatches 166; Indels 38; Gaps 5;

QY 1 PGITTAIEQAALAAANALANAVVVPRLSVQTEIILINLQPOLVFEVRLNHPFOR 60
D 10 PGITTAIEQAALAAANALANAVVVPRLSVQTEIILINLQPOLVFEVRLNHPFOR 69
QY 61 VINHELOYCARAGCLEVGAMPRSTINDPNVLRHCFLPBVGHDVORWYSAPTRGPAAN 120
D 70 VINHELOYCARAGCLEVGAMPRSTINDPNVLRHCFLPBVGHDVORWYSAPTRGPAAN 129
QY 121 CRRSALGLPRADTYGCFDFGSCAFAPETGVALYSJLHDLMPADVAAMARHGCTRXYAA 180
D 130 CRRSALGLPRADTYGCFDFGSCAFAPETGVALYSJLHDLMPADVAAMARHGCTRXYAA 189
QY 181 LHPPEVLBPFGTYHTTSYLLIHDGDAVYVTEGDTSGAGNVHDSIRAMTRTKKIYGDH 240
D 190 LHPPEVLBPFGTYHTTSYLLIHDGDAVYVTEGDTSGAGNVHDSIRAMTRTKKIYGDH 249

```



```

OY 241 PLVIERAATGCHVLLTLTAAPERSPMPYVPRSTEVYRSIFGCGSPSLPSPACSTK 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 250 PLVIERVKGIGCHVLLTLTAAPERSPMPYVPRSTEVYRSIFGCGSPSLPSPACSTK 309
OY 301 STEHAAVPHIMDLMLFGATLDDOAFCCSRMLTYLKGISYKTVGALVANEGMNASEDAL 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 310 STEHAAVPHIMDLMLFGATLDDOAFCCSRMLTYLKGISYKTVGALVANEGMNATEDAL 369
OY 361 TAITAAVYLITICORLYLQALISKMRRLGVEAOKFTIRLYSMLEKSGRDYIPGRLO 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 370 TAITAAVYLITICORLYLQALISKMRRLGVEAOKFTIRLYSMLEKSGRDYIPGRLO 429
OY 421 FYAOCRRMLISAGHLDPRVLVFEDESVCRCRPTLKKVAGFCCEFMMLGDECTFLPEAE 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 430 FYAOCRRMLISAGHLDPRVLVFEDESVCRCRPTLKKVAGFCCEFMMLGDECTFLPEAE 489
OY 481 GLVGDHNDNEAVEGSEVDPAEPBAHLDVSGTYVAHGHOLEALYRALNVPODIAARASRLT 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 490 GLVGDHNDNEAVEGSEVDPAEPBAHLDVSGTYVAHGHOLEALYRALNVPODIAARASRLT 549
OY 541 AYVELVASEDRLECKTVLGNKTRFTTVVDCALHLEANGPEQYVLSFDSAROSKAGSHSLT 600
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 550 AYVELVASEDRLECKTVLGNKTRFTTVVDCALHLEANGPEQYVLSFDSAROSKAGSHSLT 609
OY 601 YELTPAGLOVRISNGLDCTATPPPGCAPSADGEVAAPCSALYRYNRFTQHSILTGWL 660
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 610 YELTPAGLOVRISNGLDCTATPPPGCAPSADGEVAAPCSALYRYNRFTQHSILTGWL 669
OY 661 LHPEGLLIGTPPPSPGHIMESANPFCGEGTLTYRTWST-----SGFSSDFSPPE 709
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 670 LHPEGLLIGTPPPSPGHIMESANPFCGEGTLTYRTWSTIDPFLVGLISG-HLDAAPH 728
OY 710 AABAAATATGRLPHSTRPVSDIWLPRPSEPEYDAAPVPADDPACLPBPVLTLPPTPP 769
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 729 GCPAATATGRLPHSTRPVSDIWLPRPSEPEYDAAPVPADDPACLPBPVLTLPPTPP 774
OY 770 PVHAPSIPTPPSRRLLYTPPDGAKYVAGSLFSDCDMLVNAANPCGPRPGGGLCHAFYOR 829
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 775 -----QRLHHTYPPDGAKIYVGSIFSESECTWLVNANAGHRPBGGLCHAFYOR 822
OY 830 FPEAPYPTPEIMREGIAAATLTPRPITIHAAVADPYRVEQNPRLLEAAYRETCSRRTAAYP 889
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 823 YPSPFATKFEVMDGLAAATLTPRPITIHAAVADPYRVEQNPRLLEAAYRETCSRRTAAYP 882
OY 890 LUGSGIYOVVSLSPAMENNRHPGDELYLTERAANFENAKAOYVLTTEPTARTANL 949
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 883 LUGSGIYOVVSLSPAMENNRHPGDELYLTERAANFENAKAOYVLTTEPTARTANL 942
OY 950 ALBIDATEVGRACAGCTISPGIVHOFTAGVPGSGKSRIOGQDADVVPVPTRELNSW 1009
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 943 ALBIDATEVGRACAGCTISPGIVHOFTAGVPGSGKSRIOGQDADVVPVPTRELNSW 1002
OY 1010 RRRGFAATPHTAARTYIGRRVYIDEAPSLPRHLLLLHMOASSVHLGDPNOIPAIDFE 1069
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1003 RRRGFAATPHTAARTYIGRRVYIDEAPSLPRHLLLLHMOASSVHLGDPNOIPAIDFE 1062
OY 1070 HAGLVPAIRPELAPTSWAXYTHHCPADVCELINGATPKIOTTSRVLSLFWNEPALGOKL 1129
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1063 HAGLVPAIRPELAPTSWAXYTHHCPADVCELINGATPKIOTTSRVLSLFWNEPALGOKL 1122
OY 1130 VXTQAANKANPAATYVHEAGATFETTTIATADAGLISSRAHAIVALTRHTEKCVIL 1189
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1123 VXTQAANKANPAATYVHEAGATFETTTIATADAGLISSRAHAIVALTRHTEKCVIL 1182
OY 1190 DABGLLREVGISDVIVNNEFLAGGEVGHRRSVIPKGNPDQNLGTLOAPSPCOISAYHO 1249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1183 DABGLLREVGISDVIVNNEFLAGGEVGHRRSVIPKGNPDQNLGTLOAPSPCOISAYHO 1242
OY 1250 LAEELGHRAPAAVAPRPCPELEOGLLYAMPDELTVDSDVYVLELTDIVHCRMAAPSORKA 1309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1243 LAEELGHRAPAAVAPRPCPELEOGLLYAMPDELTVDSDVYVLELTDIVHCRMAAPSORKA 1302
OY 1310 VLTSLVGRGRRTKLYEAHSDVRESLARIPITGTVQATTCCLYELVEAMVEKGDGSA 1369

```

```

DB 1303 VLTSLVGRGRRTKLYEAHSDVRESLARIPITGTVQATTCCLYELVEAMVEKGDGSA 1362
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
OY 1370 VLELDLCNRDVSIRTEFPQKXCKNFTTGETTIAHGKVGOGISAWSKTCALGFWPRAIERE 1429
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1363 VLELDLCNRDVSIRTEFPQKXCKNFTTGETTIAHGKVGOGISAWSKTCALGFWPRAIERE 1422
OY 1430 IIALLPNIFYGDAFEESVFAAASGAGSCMVFNENFSEPDSTQNNFSLGLECVNMEEG 1489
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1423 IIALLPNIFYGDAFEESVFAAASGAGSCMVFNENFSEPDSTQNNFSLGLECVNMEEG 1482
OY 1490 MPQWLRLVHLVSAMTLOAPKESLKGFMKHSGEPTLLAMTVMMNAITAHCEYRDRP 1549
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1483 MPQWLRLVHLVSAMTLOAPKESLKGFMKHSGEPTLLAMTVMMNAITAHCEYRDRP 1542
OY 1550 VAAFKGDDSVYLCSDYRQSHNAALJAGCGLKLVDRPILGYAGVVAAGLCTLPDVR 1609
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1543 VAAFKGDDSVYLCSDYRQSHNAALJAGCGLKLVDRPILGYAGVVAAGLCTLPDVR 1602
OY 1610 FAGRLSEKMMGPPPERAEQRLAVCDFLGLTVAAQVYCVYRVYGVSPGLVHNLIGML 1669
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1603 FAGRLSEKMMGPPPERAEQRLAVCDFLGLTVAAQVYCVYRVYGVSPGLVHNLIGML 1662
OY 1670 QTIADGKAHFTETIKPVLDTNSIIORVE 1698
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1663 QTIADGKAHFTETIKPVLDTNSIIORVE 1691

```

```

RESULT 3
C48547
nonstructural protein - hepatitis E virus (strain Tashkent) (fragment)
C:Species: hepatitis E virus
C:Date: 17-Feb-1994 #sequence-revision 17-Feb-1994 #text-change 20-Sep-1999
C:Accession: C48547
R:Try, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Platak, M.; F
Virus Genes 6, 173-185, 1992
A:Title: Hepatitis E virus (HEV): strain variation in the nonstructural gene region e
A:Reference number: A48547; M0ID:92271462
A:Accession: C48547
A:Molecule type: genomic RNA
A:Residues: 1-290 <FRY>
A:Cross-references: GB:10337; NID:q291457; PIDN:AAA5733.1; PID:q291458
A:Note: sequence extracted from NCBI backbone (NCBIN:104577, NCBIIP:104580)
C:Suprafamily: hepatitis E virus nonstructural protein
C:Keywords: ATP; nonstructural protein

```

```

Query Match 15.2%; Score 1368; DB 2; Length 290;
Best local similarity 85.9%; Pred. No. 1,8e-78;
Matches 249; Conservative 21; Mismatches 20; Indels 0; Gaps 0;

```

```

OY 1344 GPVQATTCCLYELVEAMVEKGDGSAVLELDLCNRDVSIRTEFPQKXCKNFTTGETTIAHGK 1403
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 GPVQATTCCLYELVEAMVEKGDGSAVLELDLCNRDVSIRTEFPQKXCKNFTTGETTIAHGK 60
OY 1404 VGGGISMSTFCALGFWPRAIEREIKELALLPNITFGDAFEESVFAAASGAGSCMVFE 1463
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 VGGGISMSTFCALGFWPRAIEREIKELALLPNITFGDAFEESVFAAASGAGSCMVFE 120
OY 1464 NDFSEPDSTQNNFSLGLECVNMEEGCPQWLILYHLVRSAMTLOAPKESLKGFMKHS 1523
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 NDFSEPDSTQNNFSLGLECVNMEEGCPQWLILYHLVRSAMTLOAPKESLKGFMKHS 180
OY 1524 EPCTLLMNTVMNAITAHCEYRDRPFAAFAEKGDSDVYLCSDYRQSHNAALJAGCGLK 1583
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 EPCTLLMNTVMNAITAHCEYRDRPFAAFAEKGDSDVYLCSDYRQSHNAALJAGCGLK 240
OY 1584 VDRPITGLIAGVVAAGLCTLPDVRPAGRLSKKMMGPPPERAEQRLAV 1633
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 VGRPITGLIAGVVAAGLCTLPDVRPAGRLSKKMMGPPPERAEQRLAV 290

```

RESULT 4

QY 1234 TLOAFPSCQISAYHQIAEELGHRAPVAAVLPCPELEGGILYMPQELTVSDSVLVFEE 1293
DB 1134 TM-----FATVPYTKQMLQNSLY-----1151
QY 1294 TDIHCRMAAPSKRAVLSTLVGRGR-----RTKLYEAHSDVRESL 1336
DB 1152 ---VHNIFLPLYSKTGETYDMQEFYDRCLPGNSFVLNDEDAVTMRJRDNEFLNQPCRLLT 1208
QY 1337 ARPIPT-----TQVQATTC-----LYELVEMWEGKDGSAVL-ELDLCNR 1378
DB 1209 SNLDPVALIKMEQNFLLPVLTACERPRIPQLLENLMAIKRNNPTDLAQTVDITNM 1268
QY 1379 DVS-----RITPEOKCKNKFTGETIANGKVG-----1405
DB 1269 SISIVNFFSSFPYREVLLDHLDCVRASSIQSFSDWFSOCPPTSAVQOLANFIDLPAD 1328
QY 1406 -----OGISAMSKTFCALGPGWFRALKEKILLAPRN-1 1438
DB 1329 TYWMHMKRQPKSLDTSIQSEYPALQTIYVHPKVNAVAGPVEKYLTTKFLSVDSKFF 1388
QY 1439 FYGDAAVEEV--FAAIVSGAGSCMVENDSEFDSIQNNFSLGECVWMECGMPQLIR 1496
DB 1389 FTRKKRPEDLOEFFSDLSHSDYELLDVSKYDQSQSFHSIEMAIWEKGLDLDILAM 1448
QY 1497 LYHLVRSAMTLOAPKESLKG--FMKKHSGEPGTLTMT-----VYNNMAIIAHCEFEDE 1548
DB 1449 MMSMGKRRITLDDFOAGITTLTIYORKSGDYTFITGNTFLIACVASMPLDKCFK-----1504
QY 1549 RAAAFKGRDSSVYLC--DYROSRNAALIAAGCLKLYDRPGLYAGVYV--APGLG 1602
DB 1505 --ASFCDGDSLYLPGLEYPDIQATANLY--WNFEAKLFKKYGYEGCGYIIHANGCI 1560
QY 1603 TLPDVYRFRGRLESEKMGPGPERAEQRLAVCDFGLGNVA 1644
DB 1561 YVPDPLKLISKLGKSL-VGEHVEEERISLIDVAHSLFNGA 1601

RESULT 6
S38480
nonstructural protein - rubella virus
C:Species: rubella virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
R:Gilliam, S.
submitted to the EMBL Data Library, March 1993
A:Description: Nucleotide sequence of the nonstructural protein genes of rubella virus.
A:Reference number: S38480
A:Accession: S38480
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-2115 <GIL>
A:Cross-References: EMBL:X72393; NID:9410507; PIDN:CAAS1087.1; PID:9410508
C:Superfamily: rubella virus nonstructural polyprotein

Query Match 4.3%; Score 390.5; DB 2; Length 2115;
Best Local Similarity 19.3%; Pred. No. 5,6e-16;
Matches 452; Conservative 205; Mismatches 736; Indels 947; Gaps 105;

QY 3 ITTAFOALAAANSAANVAVRPLSRVQTEILINLMQRLQVRRPVIWMPYQRYI 62
DB 40 VYTAOKRAIV-----AVIPRFVTOYMS-----DHP---AL 69
QY 63 HNELEQYCARAGRCLEVAHPRS-----INDPNVLIHRCFLRPVGRDYQWYS-----111
DB 70 H-AISRYTR--HWIEMG--PKALHLVLIDPSGL-----LREVARVERRWALCLHRT 118
QY 112 -----APTGRPAN-----CRSALRGIRP-----ADRYTC-FDGE 141
DB 119 ARKLATALLETSEAWHADYVC--ALRGAPSGPFYHEDVPHGRGAVADRLCLTYTPM 175
QY 142 SRCFAAETGVALYSLHDLMPADVA-----E 167

DB 176 QMCELMRTIDATLLVAVDLMPALAAHVGDMDWDLGIAMVHLDHGGCPADCGAGAGPYP 235
QY 168 AMARHGXTPLYAALHLPREVLLPRGYHTTSTYLLHIDGRAVVYTBGDSAGYNEDVSL 227
DB 236 GYTRPCTTRIYOVL---PDTAFHGRLYRCGPRLMTFRDCVAELSWEAONCHQH-----286
QY 228 RAMIRFTKIVGHPVLTEVRVAIGCHFVLLTAAPERMSMPVYPRSTREYVYRSIFGPG 287
DB 287 -----ANRAVACTLPIRHVRSLOPS-----ANRY-----311
QY 288 GSPSLFPASCTKSPFHAAPVHIMRLMFGATLDDQAFCSRLMTYLIGISYKYVGL 347
DB 312 -LPDLVHLAEVGMWRFSLPFVYFQRMLSYCTLSDAIYSSEVFKNALSHSTILAON 370
QY 348 VANEGN---ASEDALTYATAVLTICHQRYLRTQAIKGMIRGLVEHAQKFTRLYSW 404
DB 371 VLOEGWKGTCAEDALCAVAAYFR---AMQSNARLAGIMKSAKCAADSLs-----VAGW 421
QY 405 LPEKSGROYIPGRQLOFYA-----OCRMLSGFHLDRVLYFDESVPCRCRTFLKVA 458
DB 422 L-----DTIWDAIKRFSGSVPLAERMEK-----EQDAVAADF-----455
QY 459 GKFCSPRWLGOECTCELEPABGLVGDHGDNEAVEGSEVDPAERALHDVSGTYAVGHQ 518
DB 456 -----RGPLEDGGRIHLDYQ-----PKSPRPPEIAATWIVHAAS 490
QY 519 LE---ALYALNVPOD-----IAARSRLT 540
DB 491 ADRHCACAPRCVDYPRERPSAPAGPDDALIPWLEFAERRALCRMEDEALRARADTAA 550
QY 541 ATVELVASPDRECRVL-----GNKTRFTYVVDGAHLEANGPEQYLS 584
DB 551 APAAPLPPARPRY--TVLRHRAHNGPMTLDEBGGADALVLCDLQORLPKEPH---605
QY 585 FPAOSRMSGASHSLTYELTPAGIQ--VRISNGLIDCTATFPFGAPSAPEVAAFCSA 642
DB 606 -----YAAGAHMCA--QARGLOAFVRY-----VPPERPMADGG--AARAKK 643
QY 643 LYRYNRFOTRHSITGLGLMLHPESLGC---IPEPSPGHIMESANPCGEGTLTYRTWST 698
DB 644 FFRGCAMAQR-----LTGEPAVMHLPYTDGVPKLI-----ALALRTLAQ 683
QY 699 SGFSSDPS-----TPEAAPRAMATPG--LPHSTPVSIVWLPRP-----SEEF 741
DB 684 QGALALASVRLDPRGTAFANAAVTAVRAGPGQLAATSPPGD---SNRRLLTY--PD 791
QY 742 QVDAAPVPPAP--DPAGLGCPVVLTPPPPVYKPSIIPP-----SNRRLLTY--PD 791
DB 740 SDARGTPPPAPVNDPP-RQOP---SPPARPRVQDP--VPTTAEPRADRAHAELEVYERS 794
QY 792 GAKVYAGSLFEESD-----CDMLVNASNPGHRRGGGLCHAF 826
DB 795 GPPTSTKADPDSDIVESYARAAGPVHLRVLDIMDPKCKVVYVNAANEGLGAGSGCAT 854
QY 827 YQRFPAFYPTFERIMEGLA--YTLTPRP-----ITNAVAP-----DY 863
DB 855 FANATAA-----LAADCRRLAPCPIGEAVATPBGCGYTHIIHAVARRRDRDA 903
QY 864 RYBONPKRLEAAYRE-----TCSRGTAAVYPLLSGTY--QYVVSISPFAMERNHPPGE 916
DB 904 ALPEGALLERATRSIVALAARARARAVACPLIGAGVYCGMSAESIRALAAATRAAPER 963
QY 917 LYL-----TEPANMFEANKPAOPVLTITTEDTART 946
DB 964 VSLHICHPRATLTNHSVVGAGIARAVSPPTPELAS--CPAGDGPRAORSASPAP 1022
QY 947 ANALEIDATEVGRACCTIS-----PGIV-----973
DB 1023 LG-----DATAPEPRGCGCELCRYTRVTNDRAYVNLMLERDGAATSWAMRIPEVVVGP 1077
QY 974 -----HQF-----TAGVPGSGKSR-----988
DB 1078 EHLATHFPLNHSVLKPAFVYRPPKMGSGDMKRCRMQGMPOVRCITPSNAHALCTGVP 1137

QY	969	STIGQDVVVVPPRELRLN-SMRRGCAATPHTAARVITGR-----	1029
Db	1138	PRVSTRGGEIDPNTCMLRAAANAQAARACAAATSGAPCKACVGRALSEARHEDFAALS	1197
QY	1030	RVVIDEAPSLPPLLHLMORA-----	1051
Db	1198	QWMSAHADASPDTGDPDLRLMETVGCACSRWVSGSEHAPRPHILVSLHRAHPGPGV	1257
QY	1052	SSVHLL-----GDPNOI--PAI-----	1066
Db	1258	VLEVAREEGGNPGHFVCVAGVGPRVRSDPHMLTAVLPSRGGTCAATDEGLAQAYD	1317
QY	1067	DPE-----HAGLVPAIREFLAPTS--WMXVY-----	1095
Db	1318	DLEVRLLGDDAMARAALASIQRPKPGYNIRVNNMAAGAKTTRILIAAFTREDLVCCPTN	1377
QY	1096	DVCELLR-----GAPYKI-----QTSR	1113
Db	1378	ALHHEIQAKLARNIDIKMNATERRALTKRLAYRRIIDEAFTLIGECVAFVASQTIAE	1437
QY	1114	VL-----RSLF-WNEP-----	1123
Db	1438	VICVGRDQCPRHYANNCRTPVDPDPMETGSRHPTWFRPCMAARLRAGLDYDEBERTGT	1497
QY	1124	ATGOKLVYTOAKAKANPCALIVHEGACATETETITATADARL-	1167
Db	1498	FACNIMDROVDLHLASRETV--RLRHEKGTAIATYVREAOQMSVGTACIHWGRGTDVA	1555
QY	1168	IQSSRAHAIVALTZHTTEKCYILD-APGLREVGISDVIYNNFFLAGGEV-----GXH	1218
Db	1556	LALTRDLAIVLSTRASDALYHLEGGELTLRAAGLS-----AFLDAGALAELEKVPACID	1609
QY	1219	PSYIRPGNDQNLGTLQAFPPSCQISAHQIAELGHRPAVVAIVPRCEL	1271
Db	1610	KVVAEQAPPLP---PADGIPEDQVPPCPRTLELIFGRAGH---PHYADLMRYTEG	1663
QY	1272	BOGLVMPQELTIVSDSVLVEFLTDYHCRMAAPASQKRAVLSTLVGRGRTKLYEAAHSD	1331
Db	1664	EREVRYKM-----ISHLNKNHTEPGEERVLASVPCYGRAG-----DGST	1707
QY	1332	VRESIARPIP---TTPVQATIT-----CLYLVANWEKQDSAVLEL-----	1373
Db	1708	LRTAVARHPREFRQIPPRVATGVAQDAEWEMRYTLRERIDLTIVYTOYMGAAEELDRYTR	1767
QY	1374	DLCRDVSRTPEFOKKCNFR-----TGETIAHGKVGOGISAMSKPYCA	1417
Db	1768	KYPEIFAGMCTAGSLSPAPLKLATKCYDAALCPRTTECHAAQKAGLEIIPAMKEWQ	1827
QY	1418	LEGPWFARIEKILLALPNIFYGDAVEESVFAAASGAGSCVMEFDESEFSDTONNFS	1477
Db	1828	VMSPHFRAIQIILIRALRPQFLVAAGHTEPEVDAMQAHYTTNAILEVDEPTEDMQTLAT	1887
QY	1478	LGLECVWMECCMPOMLIRLHYLVS-AMIIQAKRESIKGWMKHSEF	1525
Db	1888	RDVELEISALLGLPCA--BD-----XRALAGSYCTLRBELSTETGCERTSGEP	1935
QY	1526	GTLTAMTVMMNAIITAHCEYER---DEFRYA--FKGDSVVLCSVDROSNNAA-----ALI	1575
Db	1936	ATLLHNT-----VAMCMARBMYPKGVRAAGIFQGDWVITLPE---GARNAALKKTPRAEV	1968
QY	1576	AGGGLKLIKVDY--RPIGLVAGVVABGLSTLPDVVFAFGRISLEKMMGQGPERRADQLRLAV	1633
QY	1634	CDFLRGLTNAAQCVDVVSVYGVSGVLHNLIQMLOTIADSKAIFETIKIPVLDITNSI	1693
Db	2046	LDRIRG-----VVALP-----DIVVAAAAYYDVSARVLAIVAREL	2081


```

Db 62 GFTS-----EFSISRAATP-----TP-----AAALPPADDPSPTL 93
OY 758 PGVPLVPPPPPPVHKPSIP-PPSNNRRLTYTPDGAKVYAGSLFES 803
Db 94 SAPARGEPAPGATAPARAIPATHOTAHNRRLFTYTPDGSKVAFGSLFES 140

RESULT 9
501956
hypothetical protein, 195k - turnip yellow mosaic virus
C:Species: turnip yellow mosaic virus, TYMV
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 18-Jun-1999
C:Accession: 501956
R:Morch, M.D.; Boyer, J.C.; Haenni, A.L.
Nucleic Acids Res. 16, 6157-6173, 1988.
A:Title: Overlapping open reading frames revealed by complete nucleotide sequencing of
A:Reference number: 501955; MUID:88289359
A:Accession: 501956
A>Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-1844 <MOR>
A:Cross-references: EMBL:X07441; NID:962222; PIDN:CAA30322.1; PID:962224
C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase

Query Match 3.9%; Score 349; DB 2; Length 1844;
Best Local Similarity 20.6%; Pred. No. 1.8e-13;
Matches 385; Conservative 203; Mismatches 649; Indels 634; Gaps 93;

OY 157 LHD-----LMPADVAEAMARH-GXRLYALHLPPVLLPPG-----TYHTTSLIHD 204
Db 141 MHDLMVYHPSOIMDLFRKPNLERLVSLVPPPEANLSDSFPKLYTTTTRHTLHY- 199
OY 205 GDRAVVTVEGPTSGAGYNHDSILRAMITKTI-VGDHPLVIERVAIG-CHEVILLTAAP 262
Db 200 -----VPEGHAGSYNOP-SDAHSMLKINSIRLGNHNLVTLIESWGPVSHLIGRGP 252
OY 263 EPSMPVYVPRSTEVYRSIFGP----- 286
Db 253 PPDSIQAPRLMTSDLRQEPRLDVSPFRIDALIELPQATLQQLRDLRPVAYN 312
OY 287 -----GGSPSLFPSACSTKSTFHAVPVYHMDRLMGATLDDOAFCCSRLMT 333
Db 313 ALFTYTRAVRLTISDPAFARMHSSKPDHVMNADNIQT-----ALLNV 361
OY 334 YLR-GISYKTVG-----ALVANGMNSEDALTAIXITAAVLTICHQRYLTQAIKGM 387
Db 362 PLRNNVYVHVLQSPASISLTLRQHWRL--TATAVPLISFLTL-ORFL----- 408
OY 388 RLGVENHAQ-KFITRLYSWLFKESGRDYIPGRLOFYAOCRRMLSGFILDPRVLVDESV 446
Db 409 PLPLPLAEVKSITAFRRRLYRKKE-----PHNPLDY-----FILDHRRVRYHSAI 453
OY 447 PCRCRTELKVVAGKFCCEFMRLGQECTGLEPAGLGVGHGHDNEAVGSEVDAEPANL 506
Db 454 ----- 511
OY 507 DVSQTVAVHGHLEALYALNVPODIARA-----SRLTATVELVASPDR--- 551
Db 465 KLP-----HALOKAALLLRPISPLLTATPPFRSEOKSMRLNAELISLTKFALPWOASL 519
OY 552 -----LECRVTLGNKTFRTTVVDGAHLEANGDEQYVLSFDSAROSGAGSHSLTYELTPAG 607
Db 520 VLLALSSSILLHLKLFSPPTIOAH-----DYHRLHLHESYSLQMERIP-- 564
OY 608 LOVRISNGL--DCTATPPGAPSAAPGEVAAPFCSALYRYNRRTORHSLTJGLGLMHP 664
Db 565 LSIPTRTAFLEPFTPTTSTAPDRSEASLP--PAFASL-----FVPR----- 603
OY 665 GLIGIIFPP--FSRG--HIMESANPFCGECTLVTRTWSSTSGF-----S 702
Db 604 -----PPPAASSPGADPPTTTAAAPPIEPT--ORTHONSDLALESSTSTPPPPPIRS 655

```

```

OY 703 SDFSP-----DEAAPAMAAATPGL--PHSTPPVSDIWLPPSEB----- 740
Db 656 PDMTPSAVLEPELINSRRRRPQOLPATDLEPATPPPLSIPIHDDPDSADPLMGSHLH 715
OY 741 FOYDAAPVPPAPDAGLPGVPLVTPPPPPV-----HKRS-- 775
Db 716 HSLPAPPTHPRLPSSQLPAPLPTNDPTAIGVLPPEELHPRKYPENTATFLRLSLPSNH 775
OY 776 IPPSRNRRLITYTPDGAKVYAGSLFESDCDWLVNA--SNRGHPRG-----GLCH-- 824
Db 776 LPQPTLN-CLLSAVSDQTKVSEELHWSLQTLPLDSQLSNBEFTLGLSTELTALHLY 834
OY 825 -----AFYQREPEAFPTPEITMRGLAAYTLTPPIIHAAVDYBONPKRLBAAYRET 879
Db 835 NFOATVYSDRCPLIFGSPDTIKR-----IDITHTPGPHFSFG-KRLIGSQP-- 881
OY 880 CSRRGTAYRPLG-----SGITQVYVLSFDMERNHRPGDLY---LTPRANMFEA 929
Db 882 -SARGHPSDPLIRAMKSEKVSQNY-LPES-----EAMHPTSIHAKNLISNKNKGFSG 933
OY 930 NKPAQPVLTITEDTARTANLALTEIDAATEVGRACAGTISPG-----IVHYQFTAGVGS 985
Db 934 VLSLIDYSTGORTGPTPKERITIDHLYDT-----NPKITPVYHF--AGFAGCG 981
OY 986 KRSRIQ-----GDYDVVYVPTRELNSWR--RGFAA--FTPHTAARVITGRV 1032
Db 982 KTYPIQQLKTLKFKDR-VSCPTELTEWKMTAMELHSGSWRFNWESSILKSSRLIV 1040
OY 1033 IDEAPSLPHELLHMQARSSVHL--LGD-----NOIPALDFHAGLYP 1075
Db 1041 IDELYMKRGYLDLSLADPRLLEVLIIIDPLGEGYHSQSKDSSNHRLP--ETLRILP 1097
OY 1076 AIRPELAPTSWXYTHRCPADVCELI-----RGAPKIQT-----TSHVLSLF 1119
Db 1098 YI-----DMYCNW--SYRLPCIALRLFOHSMANQVIGSVSTHDSQVLTNHSASLT 1151
OY 1120 WNEPAIGOKLVYTGAAKAAANGAITVHEAOGATFETTTIATADARGLIOSSRAHAI 1179
Db 1152 FN--SLGR-----SCTISSQGLTFCDAPIVLDYTKWLS--ANGVAL 1194
OY 1180 TR-----HTEKVIIDA--PGLREVIGSIDYVNNF 1230
Db 1195 TRRSAGVQFMGPSSYVGCTNGSSAMFSDAFNNSLIDRYFPISLEPL--KLITSLPT 1251
OY 1211 AGEVGXHRPS-----VIRPGNDONGLQIAPPS 1241
Db 1252 RGPRLNGATPSASPTNHSPPNHLRPHILPSTDRPFVYVNPPLPDQGETRLDT--HFLRP 1309
OY 1242 QOISAVHQAEEELGHRPAPVAVLPPCELEOGILLYMPDELTVSDVLYVEL-----T 1294
Db 1310 SRLPLHFDLPRAI--TPPVSTSYDP--PQAKASVY--PGEF--FDSLAFPLPHADPSTR 1363
OY 1295 DIVICRNAAPS-----QKKAIVSLVGRYGRRTKLYEAAH-----SDVESLARFI 1340
Db 1364 EILHKDSSNQFPWFDRPFLSC-----OPSSLISAKHAPNHDPDLLPASINKRL-RFR 1416
OY 1341 PTIGPVQATT-----C-----ELXYLVAEMVEKGDSAVLE 1372
Db 1417 PSDSPHOITADVVGLQLFHSILCRAYSRODNGSTVPPNPELFAICISLNEYAQLSSTOS 1476
OY 1373 LDLCNRDVS-----RITFEOKXCNKFTTGETIANGKVGQISAMSKTFCALRGP--W 1422
Db 1477 TIVANASRSDPDMRHTVYKIFAKQKHVNGSIFGSMKACQOTLALMDYILVIGPYKKY 1536
OY 1423 FRAIEKEILLALPPNIF-----YGDVSESVFAAANSVAGSCAVFENDSEEDSTQ 1473
Db 1537 QRITDN--ADRPENIYSHGKTPNQLRDMQCEHL-----TSTPKIANDYAFPOSO 1586
OY 1474 NNFSLGLECYVMECGMPOMLIRLY-HLVYSAMWLQAPKESLAKGFKMKHSGEPTLLMT 1532
Db 1587 HGESVLEALMKRLNIPSHLIQHLHKTIVSTQFGPLTCM-----RLTGEPTDYDNT 1641

```


OY 1533 VNNMIIAHCEYERDFRYAAAFKGDSDVVLCSDY--ROSRNAALIIAGGLKLVDPY----- 1586
 Db 1642 DYNLAIVITISQIDVGSCTPMV--SGDSDLT---DHPLETRHDPMSVLKRLHLKRLKELTSHP 1697
 OY 1587 -----RPIGLYAGVVVAPGLIGLPD-----VVRF--AGRLSEKMGPGPER 1625
 Db 1698 LEFGYVGPAGCIRNPLALFKCMLIAVDDDALDDRLSLYEFTTGHLLGSESLMLLPDET 1757
 OY 1626 AEQLRLAVCDF 1636
 Db 1758 HVOYOSACDF 1768

RESULT 10

SI9151
 SI9151
 hypothetical protein 221 - turnip yellow mosaic virus
 N:Contains: hypothetical protein 206
 C:Species: turnip yellow mosaic virus, TYMV
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 18-Jun-1999
 C:Accession: SI9151, SI9152
 R:Diehrer, T.W.; Bransom, K.L.
 Plant Mol. Biol. 18, 403-406, 1992
 A:Title: Genomic RNA sequence of turnip yellow mosaic virus isolate TYMC, a cDNA-based
 A:Reference number: SI9150; MUID:92119261
 A:Accession: SI9151

A:Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1-1985 <DRE>
 A:Cross-references: EMBL:X16378
 A:Accession: SI9152
 A:Status: preliminary
 A:Molecule type: genomic RNA
 A:Residues: 1-1844 <DRE>
 A:Cross-references: EMBL:X16378; NID:962218; PIDN:CAA34415.1; PID:962220
 C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase
 C:Keywords: ATP

Query Match 3.8%; Score 346; DB 2; Length 1985;
 Best Local Similarity 20.5%; Pred. No. 3.1e-13;
 Matches 380; Conservatively 210; Mismatches 660; Indels 608; Gaps 94;

OY 157 LHD-----LMPADVAMARH-GXTRLYAALHLPREVLLPPG-----TYHTSYLLIHD 204
 Db 141 MDAIMYHPSQIMDLFLKRWLEKILASLVVPRAHLSDOOFTYKLTYYTTRHILAH- 199
 OY 205 GBRVAVYEGDTSAGYNHDSILRAMIRTTKI-VGDHPLVIERVRAIG-CHEVLLITTAAP 262
 Db 200 -----VEGHGAGSYNDP-SDAHSMLRINSIRLGNHLSVTILSMGVSLLIQRGTP 252
 OY 263 EFSMPPIYP-----YPRST-----EYVRSITFG 285
 Db 253 PDDPSLQAPPTLMSDLFRSYOEPRLDVVSFRIPDAIELPQATFLQOPLRDLVPRAYVN 312
 OY 286 P-----GGSPSLFPGSASTKSTFHAIVPVHIMRCLMFGATLDDQAFCSRLMT 333
 Db 313 ALFTTRAVRITRTSDPAFVFMHSSKPDHDMVTSNANDNLDTF-----ALLNV 361
 OY 334 YLR-GISYKVTG-----ALVANEGMNASEDALTAIXTAAYLTICRORYLRTQATSKGM 387
 Db 362 PLRPVVVHVILQSPILASISLVLRQHMRL--TATAVPLISFLTL-QRL----- 408
 OY 388 RLGVEHAQ-KFTIRLYSMLEFSGRDYIPGROLOFYAOCRRMLSGFHLDPVLYFDES 446
 Db 409 PLPIPLAEVKSITARRRELYRKKE--PHNPLDV-----FHIQHRIRVYHSAI 453
 OY 447 PCRCRTEFLKVVAGKFCCEFMRLGQECOTGLBPAGELVGDHGDHNAEYESEVDPAEPAHL 506
 Db 454 -----SAVRPASPPHQ 464
 OY 507 DVSQYVAHGHOLEALYRALNVPDIAARA-----SRLTAVELVAPDR-- 551
 Db 465 KLP-----HALQKALLLLRPLISPLLTATPFPRSEOKSMLPNAELSWLTKRALPLMQASTL 519

OY 552 -----LECRVLGNKTRFTTYVDGAHLLEANGPEQVYLSDASQSGASHSTIYELTPAG 607
 Db 520 VLLAESSESLHKLFPSPPTLOAQH-----DYHRIHLHPGSSYSLOWERTP-- 564
 OY 608 LOVRISSENGL---DCTAFPPGAGSAPAGVEYARCSALYRNRPTOR-----HSLTGLM 660
 Db 565 LSIPTTAFLPPTPTTSTAIPDRSEASIP---PAPAST-----FVPPPPPAASSPGA-- 613
 OY 661 LHPGLGLIFP-PFSPGHIMESANPFCGEGTLTYRTWSTSGFSSDFSP-----P 708
 Db 614 -QPTTTAAPTPIPTPIPTQAHQNSDLESSTSTPPTPIRSPDMTPSAPVLPFEINSP 672
 OY 709 EAAAPMAATPEL--PHSTPPYSDIWLVRPPEEFQVD-----AATPPAPD 753
 Db 673 RRPPLATPDLERPAHTRPPLS---IPHODPTSDVPLMCSHLLHSRLPAPHPPLPS 728
 OY 754 PAGLGPVLPFPFPFPV-----HKPS--TPPSRNRLLYT 788
 Db 729 SOLLPAPLNDPTALGPVLPFEELPRKRYPRENTATFTRLRLSPSNHLPQPTLN-CLLSA 787
 OY 789 YPDGAKVYAGSLFESDCOMLVNA--SNPCHRPDQ-----GLCH-----AFYQREPE 832
 Db 788 VSDQTKVSEHLMESLQITLIPDSQLSNEETNTLGLSTEHLTALHLNFOATVYSDRPI 847
 OY 833 AFYPTPEFTMREGLAAYTLTPRIIHAVAPDYVEQNPKRLEAYETCSRRCTAATVPLG 892
 Db 848 LEFSPDITKR-----IDITHTTGPSPHSFG-RRLISQD--SAKHNPDDPLIR 893
 OY 893 -----SGIYQVPVSLSPDAMERNHRPDELY--LTPPAWPEANPAPVLTIED 942
 Db 894 AMKSFVSGNT-LRPS-----EAINHPTISHANLISNMKNNGDVLISLDVSTGQRT 946
 OY 943 TARTANLALEIDAAEVEGRACGCTISPG---IYHYQFACVPSSGKRSISIQ----- 992
 Db 947 GPTPKERITQIDHYLDT-----NPKCTPVVHF--AGFACGKTYPIQLLTKRL 994
 OY 993 -GDVDVVVPTRELNSMR-----RGFAA--PTPHTAARTIGRRVVIDEASLPLHLL 1045
 Db 995 FKDFR-VSCPTLETRTEWKTAMELHSGSWRENTWESSILKSSRLVLDIYKMRGYLD 1053
 OY 1046 LHMORASSVHL---LGD-----NOTPAIDFEGAGVPAIRPELATPSMX 1088
 Db 1054 LSIADPALELYITIGDPLQGEYHSQKSDSNHRPS---ELRLPLTY-----DMYCMW- 1105
 OY 1089 VTHRCPADVCELL-----RGAYPKIQ-----TSRLRSILFNNEPAIGKLVYT 1132
 Db 1106 -SYRIPQCIARLEQIHSFNAMQVIGSVSTPHDQSPVLNHSASLTEN--SLGYR-- 1158
 OY 1133 QAKAANKGATVHNAQGTFTETITATADARGLQSSRAIYALR----- 1181
 Db 1159 -----SCTISSQGTFCDPALIVDNTYKMLSS--ANGVALATRSRSGVQFMGPS 1207
 OY 1182 -----HFEKCVILDA--PGLREGISDVYNNFELAGGEVGHXP-- 1220
 Db 1208 SYVGGTNGSSAMFSAFNNLSLIMDRYFSLPQL--KLITSPILTRPKLNGATPSAS 1264
 OY 1221 -----SVIPRCNPONIGTIOAFPPSQISAYHOLAEL 1254
 Db 1265 PTHRSRPNFLPHIPLSYDRDFVANSSTLPDGPETRLDT--HFLPBRSLPLHFDLPRAI 1322
 OY 1255 GHRPAPVAAVLPCEBELDEGLLMPQELTVSDSVLVEL-----TDIVHCRMAAPS-- 1305
 Db 1323 --TPPVYSTSDP-POAKASPY-PGEF--FDSLAAPFLPAHDPSTREILHMDQSSNOP 1376
 OY 1306 -QRAVISTLVGRYGRKRLKEAAH-----SDVRESLAPRIPIGVQATTT-- 1350
 Db 1377 WFDREPSLSC-----QBSLISAKHAPNHDPITLPAISNKRL--RFRSDSHQITADY 1429
 OY 1351 -----C-----ELYELVAMVEKQDSSAVLELDLCNRDVS-- 1381
 Db 1430 VLGLQFLHSLCRAYSROPNSTVVPFNPDELFAECISINEYKOLSSKIOSTIVANASSDPDW 1489

QY 1382 ---RTFFOKKXNKFTTGETTAHCKVGOGISAMSKTCALFGR---NFRATEKITALLP 1435
Db 1490 RHTTVKIFKAKQKHNDGSLFGSKACQTLALMDHYILVYGPKKQKQIFDN---ADRP 1546
QY 1436 PNIF-----YGDAYEESVFAAAGSAGSCWGFENDESPSTONNSLGEQVME 1486
Db 1547 PNITSHGCKTPNOLRDMCOEHL-----THSTPKIANDITAFQOSOGESVLEALKMK 1599
QY 1487 ECGMPOMILIRLY-HLVASAMILQAPKESLKGFWKHSSEPTGLMTVMNNAITAHCEYF 1545
Db 1600 RLNIPIHLIOLHVKLTNVSTQFGPLTCM-----RLTGEPTGYDNDINDYNAVYISQYDV 1654
QY 1546 RDEPVAAFKGDDSVYLCSDY--ROSRNAALIIACGCLKAVDY-----1586
Db 1655 GSCPIWV-SGDDSLI---DHPLPTRHDPMSVLKRLHLRFKLELTHSHPDFCGYVGPACGI 1710
QY 1587 -RPIGLAGVYVAPGLTLPD-----VYRF--AGRLSEKMNKGPERRAEQRLAVCPF 1636
Db 1711 RNPPLATCKLMTAVDDALDRLSLYTEFTTHLGLBSLMLHPETHVYQISACPDF 1768

RESULT 11

183K protein - tomato mosaic virus (strain Ob)
J02144
N:Contains: 126K protein
C:Species: tomato mosaic virus
C:Date: 03-May-1994 #sequence, revision 03-May-1994 #text, change 16-Jun-2000
C:Accession: J02144; J02143; J02158; J02157
R:Padgett, H.S.; Beachy, R.N.
Plant Cell 5, 577-586, 1993
A:Title: Analysis of a tobacco mosaic virus strain capable of overcoming N gene-mediated
A:Reference number: J02143; MID:93299124
A:Molecule type: mRNA
A:Residues: 1-1616 <PAD>
A:Cross-references: GB:111665
A:Note: this form is translated based on a read-through of the codon TAG for residue 1116
A:Accession: J02143
A:Molecule type: mRNA
A:Residues: 1-1115 <PA2>
A:Cross-references: GB:111665
A:Note: this form is translated based on the interpretation of the TAG at position 1116
R:Ikeda, R.; Watanabe, E.; Watanabe, Y.; Okada, Y.
J. Gen. Virol. 74, 1939-1944, 1993
A:Title: Nucleotide sequence of tobamovirus Ob which can spread systemically in N gene
A:Reference number: J02157; MID:93389450
A:Accession: J02158
A:Molecule type: genomic RNA
A:Residues: 1-152, 'K', 154-872, 'N', 874-1616 <IKR>
A:Cross-references: GB:D13438; MID:9436229; PIDN:BAA02700.1; PID:9436231
A:Note: the codon TAG for residue 1116 is translated to Trp
A:Accession: J02157
A:Molecule type: genomic RNA
A:Residues: 1-152, 'K', 154-872, 'N', 874-1115 <IKR2>
A:Cross-references: DDBJ:D13438; MID:9436229; PIDN:BAA02701.1; PID:9436230
C:Comment: This protein is involved in replication of the RNA genome.
F:1-1616/Product: 183K protein #status predicted <MAN1>
F:1-1115/Product: 126K protein #status predicted <MAN2>

Query Match 3.6%; Score 324; DB 1; Length 1616;
Best Local Similarity 18.0%; Pred. No. 5,6e-12;
Matches 337; Conservative 221; Mismatches 603; Indels 714; Gaps 75;

QY 94 LHCFC---LRPGRDQRYWYSAITRGPAANCRRSAL-----RGLPADRTYCFDG 140
Db 123 VHCMMRLDLDRIMRHNQDSVATYISRLNARKKVLPAQOEAFQYRSRSEVVCNNT 182
QY 141 FSRCC---AFAAETGVALYSLHLM--PAD-VAEAMARHGXTRLYALHLPEVLLPGTY 194
Db 183 FCCCESRRYSSGGRYVAISLHSLYDIPADDELGAALLRKNVHTLYAAHFHELLLEVSTV 242

QY 195 H--TTSYLLHDDGRAVYVEGDTSGAGYNDHVSILRAMIRT-----KIVGHPHYIER 246
Db 243 ELPTIGIFISFGDGIKINFCSSNESTLNSHYSMLKYCKTYEPASNRRFVYMKFILT 302
QY 247 VRAIGCHEVLLLTAPRPSMPYPPYPRSTRSEVYVRSIFGGGSSSLPSPSCST----KST 302
Db 303 VNTWFCFKYKIDT-----YTLV-----RGYHHGCDQOEYSAMEDAMHYKKT 345
QY 303 FHAV-----PVHMDRLM-----LEGATLDDO-----324
Db 346 LAMNSERIUEHDSNVNFPKMKDVIYPLFPVSLFETQKRTKEYIVSKDFYTYVLNH 405
QY 325 -----AFCCSRIMTYLRKGISYKVTVALVANEQNA-----S 356
Db 406 IRTYQAKALTYNNVLSEVESIRSHVINGVTARESDVDKALLQSMAMTFELITKLSMLK 465
QY 357 EDALTAITAYILTI-----CHOY-----LFTQAISSKMRRLGVEHAKFT 398
Db 466 DELLVSKEFTLSASVHEHVDEIKRGGMNFPSLKESLIRKKLISGSAELEIEVPMYV 525
QY 399 TRLYSMLEKSGRDYIPRQL-OEYACRRMLSGFHLDPRLVYFDESVP-----C 448
Db 526 TENDRFAETKASVEMPTIDISKDLSAEASYALSLSL-----VLENSKDFLEKTSRMC 581
QY 449 RCRTFLKRVAGKPCCFMRWLGOBCTC--FLERAEGLVGDHNDNAYE---GSEVDP 501
Db 582 AINCVPNDIAAKIYAV--LSNESGVTLPKPEPTGMAEAMKSGEKDEVILTGSCQDNT 639
QY 502 EPHALDVSGTYAAVHGHQLEALRYALNVPDITARASLTUTVELVASPDLECRTVGNK 561
Db 640 -----DLTSKMWISGSLPLCGIASBISCDTFRNE 670
QY 562 TFRFTVVDGAHLBANGPEQYVLSFDSRQSMGAGSHSLTYELTPAGLOVRISSNGLDCTA 621
Db 671 -----ELNSLEEHML--AAESYISKMSIYVS---GPLYQOQOMONYDLSLA 713
QY 622 TFPFGGAPSAPEV-----AAPCSALYRKNRTQHSITGLMLHPEGLGTFPPF 673
Db 714 -----ASLSATVSNLKLKLVKSSVGFQDLSLKGVEFVRKK---MWL-----IKPT 756
QY 674 SPGHIMESANPCGEGTLYTRWTSQSFSSDFPPEAAAMATPGLPHSTPVSQI-W 752
Db 757 LKNHSMKVQKFGKCFLLALST-----HNEPLICADAW 790
QY 733 VLPPSEEFQVDAPVAPADPAGLGPVLTTPPPPVHKPSIPPSRRRLLYTPDG 792
Db 791 SKVAVSNESMV-----YSDM 805
QY 793 AKVYAGSLFESDCDWLVNASNPQHRRGGGLCHAFYORFPEAFYPTFTIMREGLAAYTLTP 852
Db 806 AKL-----808
QY 853 RPIIHAVPDRVEQNPKRLEAVRETCSRGAATYPLGSGIYOVVSLSPAMERNHR 912
Db 809 -----RVLKSTIGEMTISVS-----823
QY 913 PGDELYLTPPANWFPAKPAQVLTITTEDTARTANLALTEIDATEVGRACAGCTISPGI 972
Db 824 -----SAKVTYLV-----831
QY 973 VHYQTAGVPGSKRSIQOG---DVDVYVVPTRBELNSRRRGFAFTPHTAARYTIG- 1028
Db 832 -----GVPGCGKTKEILRNVSEDLVLPVPEAAAMIRKRA-----NOSGKIYANN 878
QY 1029 -----RRVVIDEAPSLPPLL--TLHMORASSVYHLGDPMQIIP 1065
Db 879 DNWKTYDSFLMNLGKPGVCOFKRLFYDEGLMLHPCGYFLVYKLSLSCNEAFVSGDIOIIP 938
QY 1066 IDF-----EHAGLVPAIRBELAPTSMWVXTHRCADVCCELIRGAVP-KIQTTSRVLS 1117
Db 939 INRVQNEPFPQHSKLIIVDETEKRT-----TLRCVDVYTHLNGCYDGAVTTTSQTS 993
QY 1118 LFW-----NEPAIGOKLYXTQAAK-----AANDGAITVHEDAGATFTETTTII- 1159

A:Cross-references: EMBL:229370; NID:9488713; PIDN:CAA82559.1; PID:9619908
A:Experimental source: tobamovirus infecting cruciferae plants (cr-TMV)
A:Note: readthrough of the terminator UGA occurs between codons CAA for 1107-Gln and CAA
A:Note: the internal stop codon is translated as X
C:Superfamily: cucumber mosaic virus RNA 1 protein
F:1-1601/Product: 178K protein status predicted <PRO2>
F:1-1107/Product: 122K protein status predicted <PRO1>

Query Match 3.5%; Score 314; DB 2; Length 1601;
Best Local Similarity 21.5%; Pred. No. 2.3e-11;
Matches 204; Conservative 149; Mismatches 338; Indels 256; Gaps 48;

893 SGUYO-----PVSLSFDM-----ERNHRGDELYLERANFEANKPAQVLTIT 940
Db 736 SGWMDVRCGRWMLIKPNAKS-HAWGVAEDANHK-----LVYVLLN-DDBKP-----VC 781
Qy 941 EDARTANLALETDAATEVGRACGCTISPG-----IVHYOPTAGVPGSGKRSIQ 991
Db 782 DERWFRVANASDLSIYSDMGKLTLTCTSPNGRPPERNKAVI---LVDCVPGCGKTKETI 838
Qy 992 QG---DVVVVVVTRRLNSMRGRFPA-----FTPHTAARTIGRRVVI 1033
Db 839 EKVNFEEDLIVPGKASKMIIRANHAGYIRADKDNVSTVDSFLMHPSRVP--KRLFI 896
Qy 1034 DEAPSLPHLLH-----MORASSVHLGDPNOIPAI---DF---EHAGLVPAI 1077
Db 897 DEG-----LMHTGCNVFLLLSODVAAYVGDQOIPETICRVANFPYPAHFAKLVA 949
Qy 1078 RPLAPATSMWXYTHRCPADVCCELIRGAYP-KIOTSRVLSLFWNE-----PA 1124
Db 950 EKRYR-----RTLRCPADVTYFLNKKYDCAVACTSAVESVAEYVKGALNPITLPL 1004
Qy 1125 ICGKLVYQAAK-----AANPGAIVHEAGAGFTETTTI-ATADARGLIOSSRAHAIVA 1178
Db 1005 ECKILFTFOADKPELEKGYKDVNVEVOGETEKTALVRLSTPLEITISRASPHYVA 1064
Qy 1179 LPHHTKCK---YILD-----AGGLREVCISIVYI-NNFELA 1211
Db 1065 LPHHTKCKYVTVVLDPMNVNISEMEKLSNELLDMYRVENGIOXOLDIVAFKGTMLFVO 1124
Qy 1212 GGEVGYHR-----PSVIPGPN-----DONLGLTCAFPSPCOI-SAYHOLAEL 1254
Db 1125 TPSSGMRDMQFYNITLPL-GNSTIINEDAYTMNLRLDISLNKCDRIDPSKVQLPKE- 1182
Qy 1255 GHRPAVY-----AAVLPPCELEOGLLYM-POELTVSDSVLVELJDIVHCRMAAPS 1305
Db 1183 ---QPIFLKPKIRTAAMPRTAGLLENLVAMIKRMNNAPDLGTIDIED----- 1228
Qy 1306 QKRAVISTLVGR--GRRTKLYEAH--SDVRESLARFI-----PTIGPVQ-----AT 1349
Db 1229 ---TASLVYERKFMDSYIDKEFSGTENMTTRSFMSWLSKQSSSYQGLADENFVDLPA 1284
Qy 1350 TCELYELVEAMVAKGODGSVLELDLCNRDVSRTFFQCKCNKFTGETIHAHKVGGIS 1409
Db 1285 VDEYKIMITSQPKQ-----KLDLSIQD-----EYRALQITVY----- 1316
Qy 1410 AMSKTCALGEPWFAIEKEIALLRPN--IFYGDAYEESV--FAAASGAGSCWFEEND 1465
Db 1317 -HSKKINATJFGPMFSELTRELLERIDSKFLPYTRKTPAQIEDFFSOLDSTQAMEILFD 1375
Qy 1466 FSEFDSQNNFSLGLGECVWEECGMPOMLRLYLHLVRSAMTLOAPRESLKF----- 1517
Db 1376 ISKYDSSQNEFHCAVEKKIMKLGIDEMLAEV-----WKQGHRTKTLTDYAGIKYCTL 1428
Qy 1518 -RKHSGEPECTLLMTVMNNAITAHCEYEFDFVAAAFKGDSDVVLQDSQSNAAALIA 1576
Db 1429 WYORKSGDVTFTIGNITIIACLSMIPMDKVIKAAFCGDSLSIYP--KGLDLPDICA 1485
Qy 1577 GCGL---KLKVDYRPIGLYAG--VVVAPGLGTLPL--DVNFAGRLSKMNPGPERAEOL 1629
Db 1486 GANLMMNFKAELPRKKYGYFCGRYVIHHDGALIVYVDPLKLSKLGCKHI-RDVVHLEEL 1544

Qy 1630 RLAVCEFLKGLTNVAQCV--DVVSRVYGVSPGLVNLIGMLQTIAD 1674
Db 1545 RESLDCVANSNLNCAYFSOLDEAVAEHHTVAGSFAFCSITIKYLSD 1591

RESULT 15

NMTPV

183K protein - pepper mild mottle virus (strain Spain)

M:contains: 126K protein

C:Species: pepper mild mottle virus

C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 08-Apr-1994

C:Accession: J01312

R:Alonso, E.; Garcia-Lange, I.; de la Cruz, A.; Wicke, B.; Avila-Rincon, M.J.; Serra, J. Gen. Virol. 72, 2875-2884, 1991

A:Title: Nucleotide sequence of the genomic RNA of pepper mild mottle virus, a resist

A:Reference number: J01312; MOID:92113528

A:Accession: J01312

A:Molecule type: genomic RNA

A:Residues: 1-1611 <AIO>

A:Cross-references: GB:M81413

A:Note: readthrough of the terminator UAG occurs between the codons CAA for 1117-Gln

C:Comment: This protein may have RNA polymerase activity.

C:Superfamily: cucumber mosaic virus RNA 1 protein

F:1-1117/Product: 126K protein status predicted <PRO>

Query Match 3.5%; Score 313; DB 1; Length 1611;
Best Local Similarity 18.7%; Pred. No. 2.7e-11;
Matches 341; Conservative 227; Mismatches 642; Indels 610; Gaps 79;

Qy 94 LHRCELPVRGROVRSATRRPANCRSAL-----RGLPADRTYCPD--- 139
Db 123 VHCMPNMDLRDVMR-----HNAQDSITELYSKLAOKKKVTPYQKP-CEDKYT 171
Qy 140 -----GFSKCAFAETGVALYSLHDLMPAD-VAEAMARHGXTRLYAALH 182
Db 172 DDQSVWVSCKPFQHCBSVSHCTDKV-YAVALHSLYLD-PDEFGALLRRNVHVCYAAFH 229
Qy 183 LPPEVLLPCTGYHTSYLIHD-----GDAVVYEGDTSAGYNHDSIIRAMT--- 231
Db 230 FSENLIL-----EDSVSISDDIGAFESREGDMLNFSFVSESTLNTHSYSLNLYCKCT 283
Qy 232 ---RTKIVGDHPLVIERVRAIGCHFVLLLTAPRSPMNVYVYPRSTEVYVSISGPG 288
Db 284 YFPASSREYVMEFLVTRVMTFCFSRL-----DFEVLVGVYHNGV 326
Qy 289 SPSLFSPACST---KSTF-----HAYVHIWDRM-----LFGATLDDQ-- 324
Db 327 DKQGFYSAMDAMHKKTLAMMNSERTLEDDSSSVNWFPMKMDVYLPFEDVSLONEK 386
Qy 325 -----AFCCSRIMTYLRGISYKVTYGVALVANEGNASE 357
Db 387 RLARKEVWVKDPVYTVLNIIRTYOSKALTYANVLSEVISRSRVITNGVTASRMDV-D 445
Qy 358 DALTAHTAYLYLTICHRVYLRQAISGMRLRGEVNAQKITLILYSLEFKSGRDYIPGR 417
Db 446 KALLQISMFTFLQTKLAMIKDLY-----VQKE-----QVHSKSLTEYV-- 485
Qy 418 QLOFYAQCRW--LSAGFHLDPRVLVFEDESVPQRCFTFLKVGKFCFCFRMLGOECTCF 475
Db 486 -----WDEITAFH-----NCFPTKE-----RLNKKLITY 512
Qy 476 LEPA-EGLVGD-----HGHNEAVEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNP 529
Db 513 SEKALEIKYVDLVYTFHRLVYKEKSSVEMPV---LDVKKSL---EEAEVVMNAL--- 561
Qy 530 QDIAARSRITAVEVLAASDRLECRVTYGNKTFRTTYVDGAIHLEANGPQOYVLSPASR 589
Db 562 -----SEISLIKSDSKFD-----VDVFSRMC 582
Qy 590 QSMGAGSHSLYTELTAPGLQVIRISSNGLDCTATFPFGAGAPSAAGVAAFCALYRYNR 649
Db 583 NTLGVN-----PLVAAKVVAVAVSNESGLTLTFE---RTEANVVALA----- 621

```

QY 650 TORHSLGGLMLHPGLGIFPPSPGHIMESANFCGEGTLTYRTWSTSGFSSDSPPE 709
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 622 -----LQPTITSKEBSLKI-----VSSDVGESSLIKE 648

QY 710 AAAPAMATPCLPHSTPPVSDIWLPPSEFOYDAAPVPPAPBAGLPGVLTLPPPP 769
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 649 VVRKSEISMGLTGMT--VSDEFORSTEIESIQ----- 679

QY 770 PVHKPSIPPPSRNRRLTYTPDGAKVYAGSLFESDC--DMLVNASNPHRPGGGLCHA 825
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 680 QFHWVSTFTIRKQM-----HAMVYTGPLKVOQCKRYLDSLV----- 716

QY 826 FYQRPPEAFYPTFEMREGLAAYTLPPRIHAAVADYVEQNPRLKLEAAYRETCRRGT 885
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 717 -----ASLSAAVSNLKKIKDPA--AIDLETKERKGVY-DVCLKKWL 755

QY 886 A-----AYPLLSGIQYQVPSLSFDAMERNHRGDELITPERANWFANKPAQPYLT 938
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 756 VKPLSKGHAMGVMDSDYKCFVAL-----LTYDGEN-----I 787

QY 939 ITEDTARTANLALIDATEVGRACAGCTI----SPGIHYQFTA--GVPSGKSRSIQ 992
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 788 VCGETWRVVAASSESIYVSDMKIRATRSVLKDEPHISSAKVLYDGVPGCGKTKELIS 847

QY 993 G---DUDVYVVPTELNSMRR--GFAAFTPHTAARV-----TIGR-----RVYID 1034
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 848 RVNDEDELVLVPGQAAEMIRRRANSGLIATKENVTVDSFLMNYGRGPCQYKRLFLD 907

QY 1035 EAPSLPHLL--LLHMQRASSVHLLDPNOIPATD---FEHAGLVPAIRPELAPTSWMX 1088
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 908 EGLMLHRCGVNPLVGMSLCSFAFYGDQOIPYINRVATFPYKHLISOLEVDAVETR--R 965

QY 1089 VTHRCPADVCELIRGAYP-KIOTTSRVLSLFW-----NEPAIQKLVXTOAA 1135
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 966 TTLRCPADITFEFLNOKYEGQWCTSVTSVSHVYIOGAAMNPYSKPLKGVITFTQSD 1025

QY 1136 KA-----ANRGATVHEQOGTFTETITII-ATADARGLQSSRAHAIVALTREITEK-- 1185
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1026 KSLLSRQYEDVHTVHEQGETFEDVSLVRLTPPVGLISKOSPHLIVLSRHTRSIKYY 1085

QY 1186 CVILDA--PGLIREYG--ISDVIVN-----NPLAGGEVGHRRPSVI 1223
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1086 TVVIDAVVSVIRDLKCYSSYLLDMKYDVSTQOQIESVYKGVNLFVAAPKGTG----- 1138

QY 1224 PRGNPDNLGTLQAFPPSCQ-----ISAVHQAELGHRPAPVAAVLPCCPELEQGLLY 1277
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1139 -----DVSDMQYYDKLPNGNSTILNEYDAVTMOI----- 1168

QY 1278 MPQELTVSDSVLVE-----ELFDIVHCRMAAPSQRK-AVLSTLYGRYGR-- 1321
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1169 RENSINWKCVDLMSKSVPLPRESETTLKPIRTAAEKPRKPGLENLVAMIKRNFNSPE 1228

QY 1322 -----TKYLEAHSVDRESLARFIPTIGPVQATCELYELVEAMVERGOD 1366
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1229 LVGVVDIEDTASLVVDKFFD-AYLKEKKKKKNIPILSRAS-----LERWIEK-QE 1277

QY 1367 GSAYLEL-----DLGNRDVSRIEFQKXCNKFTTGETIAHKVQOGISAMSKTECALFG 1420
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1278 KSTIGQLADFDFIDLPAVDQYRHMIKQPKORLDLSIOTETEPAL-QTIVYHSKKINALLFG 1336

QY 1421 PMFRAIEKEILLALPPN--IFYDAYEESV--FAAAVSGAGSCWVFENDSEFDSTONNF 1476
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1337 PVSELTROLLETIDSSRFMYRTKPTQIIEFPSSDLSNVPMIDLEIDISKYDKSONEF 1396

QY 1477 SLGLECVMECCGMPQWLIRYHLVRSAMILQAPKESILKF-----WKXSGEGCT 1527
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1397 HCAVEYIWKRLGLDPLAEY-----WKHGRKTKLKDYYTAGIKTCLWYORKSGDVTT 1449

QY 1528 LMMNTVWMAAIIACHYEFRRFVNAFAKGDSDSVLC--SDYRQSRNAALIAAGCLKLKV 1584
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1450 FIGNTIIIACTLSMLPMEHLIKGAFCGDSDSILYFPKGTDFPDIIQGANLL--WNEFAKL 1507

```

```

QY 1585 DYRPIGLYAGVVA--PGIGTLPDVVAFAGRLSEKMMGSPGERAEOLRLAVCDFELRGLT 1641
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1508 FRKRYGFCGRYIIHHDRGCIYVYDPLKLSKLGAKHT-KNREHLEERTSLCDVAGSLN 1566

QY 1642 NVAQV--CVDVYSRVYGVSP 1659
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1567 NCAYYTHLNDAVGEVYKTAAP 1586

```

Search completed: May 30, 2001, 16:10:33
Job time: 215 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 16:06:23 ; Search time 20.39 Seconds

(without alignments)
1599.806 Million cell updates/sec

Title: US-09-468-147-91

Sequence: 1 PGITTAIEQAAIAAANSALA.....FTETIKPVLDITNIIQRIE 1698

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep: *
2: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep: *
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *
6: /cgn2_6/ptodata/2/1aa/6B.COMB.pep: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7448	82.6	1693	4	US-08-478-507-7
2	7439	82.5	1693	3	US-08-840-316-1
3	7439	82.5	1693	4	US-08-809-523-1
4	7439	82.5	1693	5	PCT-US93-08849A-1
5	7439	82.5	1693	5	PCT-US93-08849-1
6	2057	22.8	431	4	US-08-478-507-2
7	515	5.7	1704	4	US-08-485-355B-40
8	374.5	4.2	2205	1	US-08-093-453B-2
9	365	4.0	78	2	US-07-876-941A-31
10	233	2.6	2500	2	US-08-801-263A-2
11	233	2.6	2500	3	US-09-102-248-2
12	218	2.4	2512	2	US-08-801-263A-9
13	218	2.4	2512	3	US-09-102-248-9
14	212.5	2.4	2517	2	US-08-801-263A-5
15	212.5	2.4	2517	3	US-09-102-248-5
16	211.5	2.3	2161	3	US-09-081-320-3
17	200	2.2	2431	1	US-07-920-281C-2
18	200	2.2	2431	4	US-08-466-277-2
19	184	2.0	1390	2	US-08-770-544-2
20	178	2.0	34	2	US-07-876-941A-32
21	165.5	1.8	255	3	US-09-081-320-14
22	162	1.8	1456	1	US-08-803-973-2
23	162	1.8	1456	1	US-08-803-972-2
24	152.5	1.7	459	4	US-09-080-983-5
25	148.5	1.6	3567	2	US-07-642-734C-4
26	148.5	1.6	3567	3	US-08-439-009A-4
27	145	1.6	1463	1	US-08-157-005-3

28	145	1.6	1463	4	US-08-747-863-3	Sequence 3, Appl
29	145	1.6	3724	2	US-08-804-227C-10	Sequence 10, Appl
30	145	1.6	3724	2	US-08-804-198-4	Sequence 4, Appl
31	143.5	1.6	1315	4	US-08-899-595-3	Sequence 3, Appl
32	140.5	1.6	1255	2	US-09-080-897-4	Sequence 4, Appl
33	140.5	1.6	1255	4	US-08-899-595-1	Sequence 1, Appl
34	140.5	1.6	1255	4	US-09-323-735-4	Sequence 4, Appl
35	140	1.6	2152	4	US-09-036-987A-3	Sequence 3, Appl
36	139	1.5	902	1	US-08-396-479B-6	Sequence 6, Appl
37	139	1.5	902	1	US-08-818-823-6	Sequence 2, Appl
38	136.5	1.5	921	1	US-08-396-479B-2	Sequence 2, Appl
39	136.5	1.5	921	1	US-08-818-823-2	Sequence 2, Appl
40	136	1.5	7257	4	US-09-335-409-5	Sequence 5, Appl
41	135.5	1.5	365	4	US-09-113-309-2	Sequence 2, Appl
42	134.5	1.5	4472	2	US-08-804-227C-2	Sequence 2, Appl
43	133.5	1.5	4302	3	US-08-658-136-5	Sequence 5, Appl
44	132	1.5	1248	2	US-09-080-897-2	Sequence 2, Appl
45	132	1.5	1248	4	US-09-323-735-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-478-507-7
Sequence 7, Application US/08478507
Patent No. 6120988
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R
APPLICANT: Yarbrough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Krawczynski, Krzysztof Z
APPLICANT: Tam, Albert
APPLICANT: Fry, Kirk E
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
TITLE OF INVENTION: No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,507
FILING DATE: 07-JUN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988


```

: ATTORNEY/AGENT INFORMATION:
: NAME: Sholtz, Charles K.
: REGISTRATION NUMBER: 38,615
: REFERENCE/DOCKET NUMBER: 4600-0183.22
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 324-0880
: TELEFAX: (650) 324-0960
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1693 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-478-507-7

Query Match      82.6%  Score 7448:  DB 4;  Length 1693;
Best Local Similarity 81.8%;  Pred. No. 0;
Matches 1396;  Conservative 110;  Mismatches 169;  Indels 32;  Gaps 6;

QY 1 PGITTAIEQAALAAANSALANAVVVRPELSNVQTEIILINLMQPRQLVFEVLNNHP10R 60
D 10 PGITTAIEQAALAAANSALANAVVVRPELSNVQTEIILINLMQPRQLVFEVLNNHP10R 69
QY 61 VINHELEQYCARAGRCLENGAMPSTINDNPVLHRCFLRPGVDQVQWTSAPTRGPAN 120
D 70 VINHELEQYCARAGRCLENGAMPSTINDNPVLHRCFLRPGVDQVQWTSAPTRGPAN 129
QY 121 CRRSALRGLPRADTYCFEDGSCFAETGVALYSLHDMAPVAEAMRHGXTRLYAA 180
D 130 CRRSALRGLPRADTYCFEDGSCFAETGVALYSLHDMAPVAEAMRHGXTRLYAA 189
QY 181 LHLPEVLLRPGTYHTTSYLLHNDGRAVVTYEGDTSGAGYNHQVSLRAMTRTKIYGDH 240
D 190 LHLPEVLLRPGTYHTTSYLLHNDGRAVVTYEGDTSGAGYNHQVSLRAMTRTKIYGDH 249
QY 241 PLVIERRAIGCHVLLLTAPRSPMPYVPRSTEVYVNSITGPGSGPSELFSSACSTK 300
D 250 PLVIERRAIGCHVLLLTAPRSPMPYVPRSTEVYVNSITGPGSGPSELFSSACSTK 309
QY 301 STEFAVNVHIMDRMLFGATLDQAFCCSRMLTYLRGISYVTVGALVANGMNASDAL 360
D 310 STEFAVNVHIMDRMLFGATLDQAFCCSRMLTYLRGISYVTVGALVANGMNASDAL 369
QY 361 TAXITAAVLTICHOYRLTQAIKSGMRLLGEVNAOKFTIRLYSWLFKESGRDYIPGROLO 420
D 370 TAXITAAVLTICHOYRLTQAIKSGMRLLGEVNAOKFTIRLYSWLFKESGRDYIPGROLO 429
QY 421 FYACCRRLWSLGFHLDPRVLVFEDESVCRCRTFLKAYAGRFCCCMRNULGOCCTCLEPAE 480
D 430 FYACCRRLWSLGFHLDPRVLVFEDESVCRCRTFLKAYAGRFCCCMRNULGOCCTCLEPAE 489
QY 481 GLYGDHNDNEAYEGSEVDPAPAHLDVSGTYAVAHQLEALYRALVNPQDIAARSLRT 540
D 490 GAVGDGHDNEAYEGSEVDPAPAHLDVSGTYAVAHQLEALYRALVNPQDIAARSLRT 549
QY 541 ATVELVASPDRLCECRYLGNKTFRTTVVNGAHFANGPEQVYLSFYDASKOSMGASHST 600
D 550 ATVELVASPDRLCECRYLGNKTFRTTVVNGAHFANGPEQVYLSFYDASKOSMGASHST 609
QY 601 YELTPAGLQVRISSNGIDCTATPPRGAPSAAPGEVAAPFSAALYRYRFPQRHSLTGGLW 660
D 610 YELTPAGLQVRISSNGIDCTATPPRGAPSAAPGEVAAPFSAALYRYRFPQRHSLTGGLW 669
QY 661 LHPBGLGIFPPSPGHIWESANPFCGEGTLTYRTWS-TSGFSSDFSP-----PEAAPA 714
D 670 FHPBGLGIFPPSPGHIWESANPFCGEGTLTYRTWS-TSGFSSDFSP-----PEAAPA 729
QY 715 MAATPGILPHSTPRVSDIWLVPSESEFOVDAAPV-PPAPPAPGL-GRVVLTPPPPPVH 772
D 730 MAATPGILPHSTPRVSDIWLVPSESEFOVDAAPV-PPAPPAPGL-GRVVLTPPPPPVH 773
QY 773 KPSTIP-PPSNRRLLTYTPDGAKVYAGSLFESDCDMLVNASNPGHRGGGGLCHAFYGRFP 831

```

```

D 767 APAITHOTARRHRLFTYPDGSKVFAGSLFESTCTWLVNMSNVDBRGCGGLCHAFYGRFP 826
QY 832 EAFYPTFEIMEEGIAATYLPRLPIIHAVAPDYREONPKRLEAYRRTCSRGTAAYPL 891
D 832 EAFYPTFEIMEEGIAATYLPRLPIIHAVAPDYREONPKRLEAYRRTCSRGTAAYPL 896
QY 892 GSGIYQVPSLSPDAMERNRHPGDELILTEPAAMFPANRPAQVLTITTEDTARTANLAL 951
D 892 GSGIYQVPSLSPDAMERNRHPGDELILTEPAAMFPANRPAQVLTITTEDTARTANLAL 951
QY 952 EIDATEVGRACAGCTTSPGIVHIOFTAGVPGSGKSSISIOGGDVVVVPTRELNSMR 1011
D 952 EIDATEVGRACAGCTTSPGIVHIOFTAGVPGSGKSSISIOGGDVVVVPTRELNSMR 1011
QY 947 ELDSATDVGRACACGCRVTPGVVQYQFTAGVPGSGKSSISIOADVDVVVPTRELNSMR 1006
D 947 ELDSATDVGRACACGCRVTPGVVQYQFTAGVPGSGKSSISIOADVDVVVPTRELNSMR 1006
QY 1012 RGFAPFTPHTAARVITGRRVYIDEAPSLPHLLLLHMRASSVHLGDDPNQIPAIIDEHA 1071
D 1012 RGFAPFTPHTAARVITGRRVYIDEAPSLPHLLLLHMRASSVHLGDDPNQIPAIIDEHA 1071
QY 1007 RGFAPFTPHTAARVITGRRVYIDEAPSLPHLLLLHMRASSVHLGDDPNQIPAIIDEHA 1066
D 1007 RGFAPFTPHTAARVITGRRVYIDEAPSLPHLLLLHMRASSVHLGDDPNQIPAIIDEHA 1066
QY 1072 GLVPAIRPELAPTSWNVTRHRCPADVCELIRGAYPKIQTTSRVYLSLFWNEPAIGQRLVX 1131
D 1072 GLVPAIRPELAPTSWNVTRHRCPADVCELIRGAYPKIQTTSRVYLSLFWNEPAIGQRLVX 1131
QY 1132 TOAKKANPGALITYHEOGATFTTTIATADARGLIOSSRAHAYVALTRHTEKCVILDA 1191
D 1132 TOAKKANPGALITYHEOGATFTTTIATADARGLIOSSRAHAYVALTRHTEKCVILDA 1191
QY 1127 TOAKKANPGALITYHEOGATFTTTIATADARGLIOSSRAHAYVALTRHTEKCVILDA 1186
D 1127 TOAKKANPGALITYHEOGATFTTTIATADARGLIOSSRAHAYVALTRHTEKCVILDA 1186
QY 1192 PGLIREVIGSDVIVNNFELFAGGEGXHRPSYIPRGNPNQIGLOAPRPSQOISAYHQLA 1251
D 1192 PGLIREVIGSDVIVNNFELFAGGEGXHRPSYIPRGNPNQIGLOAPRPSQOISAYHQLA 1251
QY 1187 PGLIREVIGSDVIVNNFELFAGGEGXHRPSYIPRGNPNQIGLOAPRPSQOISAYHQLA 1246
D 1187 PGLIREVIGSDVIVNNFELFAGGEGXHRPSYIPRGNPNQIGLOAPRPSQOISAYHQLA 1246
QY 1252 EELGHRPAPVAAYLPCPELEQGLLYMPOELTVSDVSLVELDILVHCRMAAPSQRKAVL 1311
D 1252 EELGHRPAPVAAYLPCPELEQGLLYMPOELTVSDVSLVELDILVHCRMAAPSQRKAVL 1311
QY 1247 EELGHRPAPVAAYLPCPELEQGLLYMPOELTVSDVSLVELDILVHCRMAAPSQRKAVL 1306
D 1247 EELGHRPAPVAAYLPCPELEQGLLYMPOELTVSDVSLVELDILVHCRMAAPSQRKAVL 1306
QY 1312 STLVGRGRRTKLYEAHNSVRESLARFIPTIGVQATTELELVANMEKQDGSAYL 1371
D 1312 STLVGRGRRTKLYEAHNSVRESLARFIPTIGVQATTELELVANMEKQDGSAYL 1371
QY 1307 STLVGRGRRTKLYEAHNSVRESLARFIPTIGVQATTELELVANMEKQDGSAYL 1366
D 1307 STLVGRGRRTKLYEAHNSVRESLARFIPTIGVQATTELELVANMEKQDGSAYL 1366
QY 1372 ELDCNDVSRITFEFOCKXKFTTGETIANGKVOGQISAMSKTFICALFGWFAIEKIL 1431
D 1372 ELDCNDVSRITFEFOCKXKFTTGETIANGKVOGQISAMSKTFICALFGWFAIEKIL 1431
QY 1367 ELDCNDVSRITFEFOCKXKFTTGETIANGKVOGQISAMSKTFICALFGWFAIEKIL 1426
D 1367 ELDCNDVSRITFEFOCKXKFTTGETIANGKVOGQISAMSKTFICALFGWFAIEKIL 1426
QY 1432 ALRPNIFYGDAAEESVFAAASGAGSCWVFNDESEFSTONNFSGLCEVYNEECGMP 1491
D 1432 ALRPNIFYGDAAEESVFAAASGAGSCWVFNDESEFSTONNFSGLCEVYNEECGMP 1491
QY 1427 ALRPNIFYGDAAEESVFAAASGAGSCWVFNDESEFSTONNFSGLCEVYNEECGMP 1486
D 1427 ALRPNIFYGDAAEESVFAAASGAGSCWVFNDESEFSTONNFSGLCEVYNEECGMP 1486
QY 1492 QMLIRLHLVRSAMIIQAPRESLKGFWKHSGEGTLLMNTVMMAIITACYERDRVA 1551
D 1492 QMLIRLHLVRSAMIIQAPRESLKGFWKHSGEGTLLMNTVMMAIITACYERDRVA 1551
QY 1487 QMLIRLHLVRSAMIIQAPRESLKGFWKHSGEGTLLMNTVMMAIITACYERDRVA 1546
D 1487 QMLIRLHLVRSAMIIQAPRESLKGFWKHSGEGTLLMNTVMMAIITACYERDRVA 1546
QY 1552 AFKGDSDVLSDBYRQSRNMAALJAGGLKLYDRPIGLYAGVVAAPGIGTLPRDVVFA 1611
D 1552 AFKGDSDVLSDBYRQSRNMAALJAGGLKLYDRPIGLYAGVVAAPGIGTLPRDVVFA 1611
QY 1547 AFKGDSDVLSDBYRQSRNMAALJAGGLKLYDRPIGLYAGVVAAPGIGTLPRDVVFA 1606
D 1547 AFKGDSDVLSDBYRQSRNMAALJAGGLKLYDRPIGLYAGVVAAPGIGTLPRDVVFA 1606
QY 1612 GRUSEKNWGGPERABOGLAVCDLRLGNVAQVYVVSRYGVSPGLVHNLIGMLQOT 1671
D 1612 GRUSEKNWGGPERABOGLAVCDLRLGNVAQVYVVSRYGVSPGLVHNLIGMLQOT 1671
QY 1607 GRILEKNWGGPERABOGLAVCDLRLGNVAQVYVVSRYGVSPGLVHNLIGMLQOT 1666
D 1607 GRILEKNWGGPERABOGLAVCDLRLGNVAQVYVVSRYGVSPGLVHNLIGMLQOT 1666
QY 1672 IADGKAHETETIKPVLDLTNSIIORVE 1698
D 1672 IADGKAHETETIKPVLDLTNSIIORVE 1698
QY 1667 VADGKAHETESVAKPVLDLTNSIIORVE 1693
D 1667 VADGKAHETESVAKPVLDLTNSIIORVE 1693

RESULT 2
US-08-840-316-1
: Sequence 1, Application US/08840316
: Patent No. 6034567
: GENERAL INFORMATION:
: APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
: APPLICANT: Tsarev, Sergei. A., and Robinson, Robin A.
: TITLE OF INVENTION: Recombinant Proteins Of
: TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their

```

1 TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
2
3 NUMBER OF SEQUENCES: 111
4
5 CORRESPONDENCE ADDRESS:
6 ADDRESSEE: MORGAN & FINNEGAN
7 STREET: 345 PARK AVENUE
8 CITY: NEW YORK
9 STATE: NEW YORK
10 COUNTRY: USA
11 ZIP: 10154
12
13 COMPUTER READABLE FORM:
14 MEDIUM TYPE: FLOPPY DISK
15
16 COMPUTER: IBM PC COMPATIBLE
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: WORDPERFECT 5.1
19
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/840,316
22 FILING DATE: 11-APR-1997
23
24 CLASSIFICATION: 424
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER:
27 FILING DATE:
28
29 CLASSIFICATION: 424
30 ATTORNEY/AGENT INFORMATION:
31 NAME: Richard W. Bork
32 REGISTRATION NUMBER: 36,459
33 REFERENCE/DOCKET NUMBER: 2026-4255
34
35 TELECOMMUNICATION INFORMATION:
36 TELEPHONE: (212) 758-4800
37 TELEFAX: (212) 751-6849
38
39 INFORMATION FOR SEQ ID NO: 1:
40 SEQUENCE CHARACTERISTICS:
41 LENGTH: 1693 AMINO ACID RESIDUES
42 TYPE: AMINO ACID
43 STRANDEDNESS: UNKNOWN
44 TOPOLOGY: UNKNOWN
45
46 US-08-840-316-1

OY 1547 DFRVAEKGDDSVVLCSDYROSRNAALIAAGCGKLKLVDRPGLYAGVVAAPGLGTLPD 1606
1542 DLGVAAFKGDSDIYLCSEYRSPGAVLIGCGGLKLVDFRPIGLYAGVVAAPGLGALPD 1601
OY 1607 VVRAGLSEKNMGPGGERAEQLRLAVCDPLRGLTNNVAVYVSRVYGVSPCLVHNL 1666
1602 VVRAGRLTEKNMGPGGERAEQLRLAVSDPLRKLTNNVAVYVSRVYGVSPCLVHNL 1661
OY 1667 GMLQFIADGKAHFEETIKPVLDTNSTIIOUYE 1698
1662 EMLQAVADGKAHFTESVYPVLDLTNSTICRYE 1693
RESULT 3
US-08-809-523-1
Sequence 1, Application US/08809523
Patent No. 6207416
GENERAL INFORMATION:
APPLICANT: Tsarev, Sergei. A., Emerson,
APPLICANT: Suzanne U., Purcell, Robert H.
TITLE OF INVENTION: Recombinant Proteins Of
TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
NUMBER OF SEQUENCES: 107
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/809,523
FILING DATE: 28-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13102
FILING DATE: 03-OCT-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US08/316,765
FILING DATE: 03-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/947,263
FILING DATE: 18-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: Richard W. Bork
REGISTRATION NUMBER: 36,459
REFERENCE/DOCKET NUMBER: 2026-4032US4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 AMINO ACID RESIDUES
TYPE: AMINO ACID
STRANDEDNESS: UNKNOWN
TOPOLOGY: UNKNOWN
US-08-809-523-1
Query Match 82.5%; Score 7439; DB 4; Length 1693;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 1397; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

DB 70 VHNLELYCARASGRCLETCAGHRSINDNPNVLRHCFRBPVGRDVOYRWYSAPTRGPAAN 129
OY 121 CRSSALRGLEPPADRYCFDFSPSCAFAEFGVALYSLHDMLPADVAAAMRHGXTRLYAA 180
130 CRNSALRGLEPPADRYCFDFSPSCAFAEFGVALYSLHDMSPDVAAAMRHGXTRLYAA 189
OY 181 LHLPEVLLPPTGYHTTSYLLIHGDRAVVTYEGDTSAGYNHDSILRAVIRTKLYGDH 240
190 LHLPEVLLPPTGYHTTSYLLIHGDRAVVTYEGDTSAGYNHDSILRAVIRTKLYGDH 249
OY 241 PLVIERVRAIGCHEVLLTLTAPEPSMPVYPRSTEVYVRSIFPGGCSLPPSACSTK 300
250 PLVIERVRAIGCHEVLLTLTAPEPSMPVYPRSTEVYVRSIFPGGCSLPPSACSTK 309
OY 301 STFAVAVHIMDRMLMGLATLDDOAFCCSRLMTYLRGISYKVMGALVANEGNASDAL 360
310 STFAVAVHIMDRMLMGLATLDDOAFCCSRLMTYLRGISYKVMGALVANEGNASDAL 369
OY 361 TAYTAAVLTICHOVRLTQAISKGMRLGVEHAKETITRLYSWLFKSGRDYIPGRLO 420
370 TAYTAAVLTICHOVRLTQAISKGMRLGVEHAKETITRLYSWLFKSGRDYIPGRLO 429
OY 421 FYACRRMLLSAGRLHDPRLVLPDESVCRCRTPLKAKAGFCCEMRNLGDECTCLEPAE 480
430 FYACRRMLLSAGRLHDPRLVLPDESVCRCRTPLKAKAGFCCEMRNLGDECTCLEPAE 489
OY 481 GLVGDHNDNEAYGSEVDPAEPAHLDDVSGTYAVHGHQLEALRALVPODIAARASLT 540
490 GLVGDHNDNEAYGSEVDPAEPAHLDDVSGTYAVHGHQLEALRALVPODIAARASLT 549
OY 541 ATVELVASPDLRECRVYLGKRTFTTVVDGAHLEANGPEQVYLSFDSAROSMGAGSHLT 600
550 ATVELVASPDLRECRVYLGKRTFTTVVDGAHLEANGPEQVYLSFDSAROSMGAGSHLT 609
OY 601 YELTPAGLOVRISSNGLDCTATPPRGAPSAARCEVAAPFCASALYRYRFTQKRLSLTGLM 660
610 YELTPAGLOVRISSNGLDCTATPPRGAPSAARCEVAAPFCASALYRYRFTQKRLSLTGLM 669
OY 661 LHPGGLGIFPPSPGHIWESANPFCGEGTLTYRTWS-----TSGFSDFSFPE 709
670 LHPGGLGIFPPSPGHIWESANPFCGEGTLTYRTWS-----TSGFSDFSFPE 724
OY 710 AAAPAMAATPGLPHSTPPVSDIWLPPSEFOVDAAPV-PAAPDPA-GLPGPVYLTTPP 767
725 PSIPSRATP-----FP-----AAPLPAPDPSPPTLSABRGPAP 761
OY 768 PPRVHKRSIP-PPSRNRRLTYTPDGAKYAGSLFESDCMVLNANSGRRPGGGLCHAF 826
762 GATAPAPATIHOTABRRRLFTYPPDGSKVPAGSLFESTCTWLVNASVNDRRPGGGLCHAF 821
OY 827 YORPEAFYPTFELIREGLAATLTPRPIIHAAVADVRVONPKRLAAARETCSRGTA 886
822 YORPEAFYPTFELIREGLAATLTPRPIIHAAVADVRVONPKRLAAARETCSRGTA 881
OY 887 AYPLGSGIYQVPSLSEFDMERNHRBGDELYLEPAAWFEANKPAQVLTJTEDTART 946
882 AYPLGSGIYQVPSLSEFDMERNHRBGDELYLEPAAWFEANKPAQVLTJTEDTART 941
OY 947 ANLALEIDAATEVGRACAGCTISPGIVHIOFTAGVPESGKSRSIOGCDVYVVPYPRLELR 1006
942 ANLALEIDAATEVGRACAGCTISPGIVHIOFTAGVPESGKSRSIOGCDVYVVPYPRLELR 1001
OY 1007 NSWRRRGFAFTPHTAARVYIGRRVYIDEAPSLRPHILLHMRASSVHLLGDPNOIPAI 1066
1002 NAMRRGFAFTPHTAARVYIGRRVYIDEAPSLRPHILLHMRASSVHLLGDPNOIPAI 1061
OY 1067 DFHAGLVPAIRBELAPTSMWXYTHRCPADVCELICGAYKPIQTSRVLSLFWNPEPAIG 1126
1062 DFHAGLVPAIRBELAPTSMWXYTHRCPADVCELICGAYKPIQTSRVLSLFWNPEPAIG 1121
OY 1127 QKLVXTQAAKANPGAITYHEAGATFETTTIATADARGLIOSSRAHAIVALTTRHTEKC 1186
1127 QKLVXTQAAKANPGAITYHEAGATFETTTIATADARGLIOSSRAHAIVALTTRHTEKC 1186

```

Db 1122 OKLVTOAKANKANGSVTVHEAGATYETETIIIAFADARGLIOSSRAHAIVAFPHTEKC 1181
Oy 1187 VILDPGLLREVGLSDVIVNNEFLAGEVGHXRDVIRGNDONLGLQAFPSCQISA 1246
Db 1182 VILDPGLLREVGLSDVIVNNEFLAGEVGHXRDVIRGNDONLGLQAFPSCQISA 1241
Oy 1247 YHOLAEELGHRPAPAAVLPCCPELEOGLLMPQELVSDSVVLELIDYHCHMAAPSO 1306
Db 1242 FHEILAEELGHRPAPAAVLPCCPELEOGLLMPQELVSDSVVLELIDYHCHMAAPSO 1301
Oy 1307 RKAVALSTLVGGRYGRRTKLYEAHSDVRESLARFIPTIGPVQATTCELYELVEMWEEKOD 1366
Db 1302 RKAVALSTLVGGRYGRRTKLYEAHSDVRESLARFIPTIGPVQATTCELYELVEMWEEKOD 1361
Oy 1367 GSAVLELDLCNRDVSRTFFQKXCKNKTGTTGETIAHGKVGOGISAMSKTFCALFGWFRAT 1426
Db 1362 GSAVLELDLCNRDVSRTFFQKXCKNKTGTTGETIAHGKVGOGISAMSKTFCALFGWFRAT 1421
Oy 1427 EKELIALIPNIFYGDAEESVFAAASGAGSCWFEFDFSEPDSTONNFSIGECVYME 1486
Db 1422 EKELIALIPNIFYGDAEESVFAAASGAGSCWFEFDFSEPDSTONNFSIGECVYME 1481
Oy 1487 ECGMPQWILIRLYHLVRSAMIIQAPKESLKGFWKHSGEPGLLMTVNMMAIIAHCEYER 1546
Db 1482 ECGMPQWILIRLYHLVRSAMIIQAPKESLKGFWKHSGEPGLLMTVNMMAIIAHCEYER 1541
Oy 1547 DFRVAAPGDDSVLCISYROSRRNAALIIAGGCLKYDVRPIGLYAGVVAAPGIGLPLD 1606
Db 1542 DLOVAAFQDDSVLCISYROSRRNAALIIAGGCLKYDVRPIGLYAGVVAAPGIGLPLD 1601
Oy 1607 VYRFAGRLSEKMMGPGPRABQLRLACDFLGLTFNAQVCDVYVSRYGVSPGLVHMLI 1666
Db 1602 VYRFAGRLSEKMMGPGPRABQLRLACDFLGLTFNAQVCDVYVSRYGVSPGLVHMLI 1661
Oy 1667 GMLQTIADGKAHFTETIKPVLDTNIIISIORVE 1698
Db 1662 EMLQAVADGKAHFTESVKPVLDTNIIISIORVE 1693

```

RESULT 4
 PCT-US93-08849A-1
 : Sequence 1, Application PC/TUS9308849A
 : GENERAL INFORMATION:
 : APPLICANT:
 : TITLE OF INVENTION: Recombinant Proteins Of
 : TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
 : NUMBER OF SEQUENCES: 98
 : CORRESPONDENCE ADDRESS:
 : ADDRESSSEE: MORGAN & FINNEGAN
 : STREET: 345 PARK AVENUE
 : CITY: NEW YORK
 : STATE: NEW YORK
 : COUNTRY: USA
 : ZIP: 10154
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: FLOPPY DISK
 : COMPUTER: IBM PC COMPATIBLE
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: WORDPERFECT 5.1
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: PCT/US93/08849A
 : FILING DATE: 17-SEP-1993
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US07/947,263
 : FILING DATE: 18-SEP-1992
 : ATTORNEY/AGENT INFORMATION:
 : NAME: William S. Feller
 : REGISTRATION NUMBER: 26,728
 : REFERENCE/DOCKET NUMBER: 2026-4032 PCT
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (212) 758-4800
 : TELEFAX: (212) 751-6849

```

: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1693 AMINO ACID RESIDUES
: TYPE: AMINO ACID
: STRANDEDNESS: UNKNOWN
: TOPOLOGY: UNKNOWN
PCT-US93-08849A-1

Query Match      82.5%; Score 7439; DB 5; Length 1693;
Best Local Similarity 81.6%; Pred. No. 0;
Matches 1397; Conservative 111; Mismatches 162; Indels 42; Gaps 7;

Oy 1 PGITTAIEQAAALAAANSALANAVYVPRFLSKYQVEILNLQPROLVPRVYLANHPFOR 60
Db 10 PGITTAIEQAAALAAANSALANAVYVPRFLSKYQVEILNLQPROLVPRVYLANHPFOR 69
Oy 61 VIHNELEQYCARARAGRCLEVGAAHPSINDNPNVLRHCFRLPRVGRDORWYSAPTRGPAAN 120
Db 70 VIHNELEQYCARARAGRCLEVGAAHPSINDNPNVLRHCFRLPRVGRDORWYSAPTRGPAAN 129
Oy 121 CRRSALRGLPPADRTYCDGFSRCAFAAETGVALYSLHDLMPADVAEAMARHGRTRLYAA 180
Db 130 CRRSALRGLPPADRTYCDGFSRCAFAAETGVALYSLHDLMPADVAEAMARHGRTRLYAA 189
Oy 181 LHLPEVLLPRTYHTTSTYLLIHGDDRAVYVEGDTSGAGYHNHVSILAAITRTKYVDH 240
Db 190 LHLPEVLLPRTYHTTSTYLLIHGDDRAVYVEGDTSGAGYHNHVSILAAITRTKYVDH 249
Oy 241 PLVIERVRAIGCHFYLLTAAPEPSPMPYVPRSTEVYVRSIFPGGSPSLFPGSCTK 300
Db 250 PLVIERVRAIGCHFYLLTAAPEPSPMPYVPRSTEVYVRSIFPGGSPSLFPGSCTK 309
Oy 301 STEHAUVHINDRLMFCATLDDQAFCCSRMTYLRGISYKVYGAALYANEGMNASDAL 360
Db 310 STEHAUVHINDRLMFCATLDDQAFCCSRMTYLRGISYKVYGAALYANEGMNASDAL 369
Oy 361 TAXITAAVLTICHOXYLTQOATISKGMRLGVHACKFTIRLYSWLFEESGROYTGRLO 420
Db 370 TAXITAAVLTICHOXYLTQOATISKGMRLGVHACKFTIRLYSWLFEESGROYTGRLO 429
Oy 421 FYAOCRRMLSGAFHLDPRVLVDESVPCRCRTFLKVGAKGCCFMRWGOECTCLEPAE 480
Db 430 FYAOCRRMLSGAFHLDPRVLVDESVPCRCRTFLKVGAKGCCFMRWGOECTCLEPAE 489
Oy 481 GLVGDHGDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPDIIAARASRLT 540
Db 490 GLVGDHGDNEAYEGSEVDPAEPAHLDVSGTYAVHGHOLEALYRALNVPDIIAARASRLT 549
Oy 541 ATVELVASPDLRECTVYLGNTKFTTYVDGAILENGPEQYVLSRDSAROSAGASHST 600
Db 550 ATVELVASPDLRECTVYLGNTKFTTYVDGAILENGPEQYVLSRDSAROSAGASHST 609
Oy 601 YELTPAGLOVRITSSNGLDCTATFPPGAPASAPAGEVAFCALYRNRFORTOSHSTJGLM 660
Db 610 YELTPAGLOVRITSSNGLDCTATFPPGAPASAPAGEVAFCALYRNRFORTOSHSTJGLM 669
Oy 661 LHPGELGIFPPSPDGHIMESANPECGEGTLYTRWS-----TSGFSSDSFSPE 709
Db 670 LHPGELGIFPPSPDGHIMESANPECGEGTLYTRWS-----TSGFSSDSFSPE 724
Oy 710 AAAPMAATPGILPHSTPPVSDIWLVPPESEEFQVNAAY-PAAPRA-GLGEPVLTLP 767
Db 725 PSIPRAATP-----TP-----AAIPPPAPDPSPPLSAPARGEP 761
Oy 768 PPPVAKPSIP-PPSRNRRLTYTPDGAKYAGSLFESCDMLVNSNPGCHRGGLGCHAF 826
Db 762 GATAPAPATHTOTAHRRRLFTYPPGSKVFPAGSLFESTCTWLYNNSNDHRGGGLCHAF 821
Oy 827 YORPEAFYPTFEIMREGILAAVTLPRPIIAHAVADYRVEONPKRLEAAYRETCSRGT 886
Db 822 YORPEAFYPTFEIMREGILAAVTLPRPIIAHAVADYRVEONPKRLEAAYRETCSRGT 881

```

QY	887	AYPLLGSIYOVAPVSEFDEAMERNHRHGGDELYLTPEAPAMFEAKRQAOVLITTEDPAT	946
Db	882	AYPLLGSIYOVAPVSPGFDEAMERNHRHGGDELYLTPEAPAMFEAKRQAOVLITTEDPAT	941
QY	947	ANLALEIDATEVGRACAGCTISPGIVHOFTAGVBSGKSRSIOGDVDVVVVTRELR	1006
Db	942	ANLALEIDDATEVGRACAGCTISPGIVHOFTAGVBSGKSRSIOGDVDVVVVTRELR	1001
QY	1007	NSMRRRGPAATPHTAARVITIGRRVVIDEAPSLPPIHLILHQRASSVHLGDPMQIPAI	1066
Db	1002	NAMRRRGPAATPHTAARVITIGRRVVIDEAPSLPPIHLILHQRATVHLGDPMQIPAI	1061
QY	1067	DEFEHAGLPAIRPELPAATSMWXYTHRCPAVCELCIGAVPKIOOTSRVLSLFEMNEPAIG	11226
Db	1062	DEFEHAGLPAIRPELPAATSMWXYTHRCPAVCELCIGAVPKMIOOTSRVLSLFEMNEPAIG	11221
QY	1127	OKLYVTOAKANPGALITVHEAGATFETTTIATADANGLIOSFAHAIVALTRHTEKC	1186
Db	1122	OKLYVTOAKANPGALITVHEAGATFETTTIATADANGLIOSFAHAIVALTRHTEKC	1181
QY	1187	VIIIDAPGLIRVGVISDVIVNNFFLAGEVGXHHRPSVITPGNPDQNLGTQIAPPPSCQISA	1246
Db	1182	VIIIDAPGLIRVGVISDVIVNNFFLAGELIGHQRPVSIPIGNPDANDTLIAAPPPSCQISA	1241
QY	1247	YHOAAEEELGHRPAPVAANVLPPCELEBGLLYMPOELTVSDSYVLEFELDIYHCRMAAPSO	1306
Db	1242	PHLEIAEELGHRPAPVAANVLPPCELEBGLLYLPEQELITDCDSVYTFELDIYHCRMAAPSO	1301
QY	1307	RKAAVLSTLVGRYGRRTKLYEBAHSDVNEISLAREIPTIGPVQATTCELYELVEAMVEKGD	1366
Db	1302	RKAAVLSTLVGRYGRRTKLYEBAHSDVNEISLAREIPTAIGPVQATTCELYELVEAMVEKGD	1361
QY	1367	GSAYLELDDLCRROVSRITTFQKCKNFTTGETIABRKVQOGISAMSKTFCALFGWPAFI	1426
Db	1362	GSAYLELDDLCRROVSRITTFQKCKNFTTGETIABRKVQOGISAMSKTFCALFGWPAFI	1421
QY	1427	EKEIALLLPPIFIYGDAYEESEVAAAASGASCMVEENFSEPDSTONNFSGLGECVME	1486
Db	1422	EKALITALLPQGVIFGDAFDDTVESAAYAAAKASAMVEENFSEPDSTONNFSGLGECALME	1481
QY	1487	ECGMPQWLIRLYHLVRSAMITLOAPKRESLKGFWKKHSGEGTLLMNTVMNAITIAHCYFR	1546
Db	1482	ECGMPQWLIRLYHLIRSAMITLOAPKESIRGFWKKHSGEGTLLMNTVMNAIVITHCYFR	1541
QY	1547	DEFAAARFGDSDSYVLCSDYBQOSNAALALAGGLKTLKDYRPIGLYAGVVAAPSGGLTDP	1606
Db	1542	DLOVAARFGDSDSYVLCSEYBQOSGAAILLAGGLKTLKDYRPIGLYAGVVAAPSGGLTDP	1601
QY	1607	VVRFAGRLSEKNMGPGPERAEQOLRLAVCDLRLGTLTVADQVADVVSRYGVSPGLVHMLI	1666
Db	1602	VVRFAGRLTEKNMGPGPERAEQOLRLAVSDFLRLKTLTVADQVADVVSRYGVSPGLVHMLI	1661
QY	1667	GMLOITADGKAHFEITKRPVLDITNSIQORVE	1698
Db	1662	EMLOADVADGKAHFEISYKRPVLDITNSILCRVE	1693

RESULT 5
PCT-US93-08849-1
Sequence 1, Application PC/TUS9308849
GENERAL INFORMATION:
APPLICANT: Tsarev, Sergei A., Emerson,
APPLICANT: Suzanne U., Purcell, Robert H.
TITLE OF INVENTION: Recombinant Proteins Of
TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
NUMBER OF SEQUENCES: 98
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA

```

1 ZIP: 10154
2
3 COMPUTER READABLE FORM:
4
5 MEDIUM TYPE: FLOPPY DISK
6
7 COMPUTER: IBM PC COMPATIBLE
8
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: WORDPERFECT 5.1
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: PCT/US93/08849
16
17 FILING DATE: 17-SEP-1993
18
19 CLASSIFICATION:
20
21 PRIOR APPLICATION DATA:
22
23 APPLICATION NUMBER: 07/947,263
24
25 FILING DATE: 18-SEP-1992
26
27 NAME:
28
29 ATTORNEY/AGENT INFORMATION:
30
31 NAME: Bork, Richard, W.
32
33 REGISTRATION NUMBER: 36,459
34
35 REFERENCE/DOCKET NUMBER: 2026-40322
36
37 TELECOMMUNICATION INFORMATION:
38
39 TELEPHONE: (212) 758-4800
40
41 TELEFAX: (212) 751-6849
42
43 INFORMATION FOR SEQ. ID NO.: 1:
44
45 SEQUENCE CHARACTERISTICS:
46
47 LENGTH: 1693 amino acid residues
48
49 TYPE: amino acid
50
51 STRANDEDNESS: unknown
52
53 TOPOLOGY: unknown
54
55 PCT-US93-08849-1

```

Query Match	82.5%	Score 7439;	DB 5;	Length 1693;
Best Local Similarity	81.6%	Pred. No. 0;		
Matches 1397;	Conservative 111;	Mismatches 162;	Indels 42;	Gaps 7

QY	1	PGTALIEOALAAANSALAAVAVRPLEISVQCEIILINMOPOLVREPVILNHPIOR	60
Db	10	PGITTAIEOALAAANSALAAVAVRPPLSHQOEIILINMOPOLVREPVFNHPIOR	69
QY	61	VINHELEOYCARAGRCLEVGAMPBRSINDNNVYLHRCFLRPVGDOVQWVSAIPRGAAN	120
Db	70	VINHELELYCARBSGRCLIEGAHPRSINDNNVYHRCFLRPAGRDVQWVTAIPRGAAN	129
QY	121	CRSALRGLPAPDRTYCEDFGSRCAFAEFGVALYSLHDLMPADVAEAMAHGXTRLYAA	186
Db	130	CRSALRGLPAPDRTYCEDFGSGCNFAETGIALYSLHDMSPSYVAEAMERHGTRLYAA	189
QY	181	LHLPEVLLPGTGHHTSYLLIHGCDPAAVYIECDTSGAGNVHDVSLIRAMIRTKIYGDH	244
Db	190	LHLPEVLLPGTGHHTSYLLIHGGRVAVYIECDTSGAGNVHDVSLIRAMIRTKIYGDH	249
QY	241	PLVIERVAICHPVLLLTFAPEPSPMPYPVPSTEVYVRSIFGPGGSPSLFPSSACSTK	300
Db	250	PLVIERVAICSHVLLLTFAPEPSPMPYPVPSTEVYVRSIFGPGGTSPLFTSOSTK	309
QY	301	STFAVPVHIMDLMLFGATLDDOAFCCSRMLTYLRGISTIKVYIGALVANDGNASDAL	360
Db	310	STFAVPVHIMDLMLFGATLDDOAFCCSRMLTYLRGISKVTYIGTIVANDGNASDAL	369
QY	361	TAXITTAAYLTICHOHYRLTQOASISGMRLGVEHNAOKITRLYSWLFKSSRDYIPGQOQ	420
Db	370	TAXITTAAYLTICHOHYRLTQOASISGMRLLEEHNAOKITRLYSWLFKSSRDYIPGQOQ	429
QY	421	FYACCRMLSGFHLDRPVLFEDSVRCRTEFLFKVAGKFCCEMRNLGEGCCELEPAE	480
Db	430	FYACCRMLSGFHLDRPVLFEDSAPCRHCTAIRKAVSFFCCPMKMLGEGCCEFLDPAE	489
QY	481	GLVGDHGHDEAVEGSEVDAPEAPAHLDVSGTYAVHGOLEBALYRALNVPODIARASRLT	540
Db	490	GLVGDQGHNDENAEVGEVDPAESAISDISGSYVVPCTALPLVQALDPLAEIVARAGRLT	549
QY	541	ATVELVASPDRLECRVYLGKKTFTTYVQCAHLEANGPEQOYVLSFDASRQSMGAGHSILT	600
Db	550	ATVELVQVDRIDDETLGLKKTFTFSVDAVAVETJNPEKHNLISFPAOSQSTMAAGPSSLT	609

Matches	381: Conservative	26: Mismatches	24: Indels	0: Gaps
QY	1257	RAPVAALVPCPELEQGLLPMQELTVSDSLVELNDIVICRMAAPSRKRAVSLTVG	1316	
Db	1	RAPVAALVPCPELEQGLLPMQELTVSDSLVELNDIVICRMAAPSRKRAVSLTVG	60	
QY	1317	RGRRTKLYEAASHSVRESLAFRTFTIGVQATCEIYELVEAMVEKGDGSAYLELDC	1376	
Db	61	RRCGRFTKLYNASHSVRSLARFRTAIGVQYTTCEIYELVEAMVEKGDGSAYLELDC	120	
QY	1377	NRDVSRITTFEQKXCKNFTTGETIANGKVGQGISAMSKTFCALFGWFRATKEKTLALPP	1436	
Db	121	NRDVSRITTFEQKXCKNFTTGETIANGKVGQGISAMSKTFCALFGWFRATKEKTLALPP	180	
QY	1437	NIFYDATEESYFAAASVAGSCGMFEENDESEFSDTONNFSIGLPCVYMEECGMQWLIR	1496	
Db	161	GVFYDADDDYFSAVAANAARASMFENDESEFSDTONNFSIGLPCVYMEECGMQWLIR	240	
QY	1497	LYHLVRSAMIIQAPKESLKGFWKKHSGEBGTLMLMTVMNMAIIAHCEYRDFRVAAFKGD	1556	
Db	241	LYHLRSAMIIQAPKESIRLGFKKHSGEBGTLMLMTVMNMAVITHCYDFRDFQVAAFKGD	300	
QY	1557	DSVYLCSSYRGRNMAALIALGGLTKLYDXRPIGLYACVYVAPGGLTLPDYVRFGRRLSE	1616	
Db	301	DSVYLCSSYRGRNMAALIALGGLTKLYDXRPIGLYACVYVAPGGLTLPDYVRFGRRLSE	360	
QY	1617	KNMGPGERAEQRLAVCDLFLGLTNVAQVCDVYVSRVYGVSPGLVHNLIGMLQITADGK	1676	
Db	361	KNMGPGERAEQRLAVSDFLRKLNTNVAQVCDVYVSRVYGVSPGLVHNLIGMLQAVADGK	420	
QY	1677	AHFTETIKPVL 1687		
Db	421	AHFTESVKPVL 431		
RESULT	7	US-08-485-355B-40		
		Sequence 40 Application US/08485355B		
		Patent No. 6177075		
		GENERAL INFORMATION:		
		APPLICANT: Christian, P. D., Gordon, K. H. J., Hanzlik, T. N.		
		TITLE OF INVENTION: Insect Viruses and Their Uses in		
		Protecting Plants		
		NUMBER OF SEQUENCES: 57		
		CORRESPONDENCE ADDRESS:		
		ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP		
		STREET: Four Embarcadero Center, Suite 3400		
		CITY: San Francisco		
		STATE: California		
		COUNTRY: United States		
		ZIP: 94111-4187		
		COMPUTER READABLE FORM:		
		MEDIUM TYPE: Floppy disk		
		COMPUTER: IBM PC compatible		
		OPERATING SYSTEM: PC-DOS/MS-DOS		
		SOFTWARE: Patent Release #1.0, Version #1.30		
		CURRENT APPLICATION DATA:		
		APPLICATION NUMBER: US/08/485,355B		
		FILING DATE: 07-Jun-1995		
		CLASSIFICATION: <unknown>		
		PRIOR APPLICATION DATA:		
		APPLICATION NUMBER: US 08/440,522		
		FILING DATE: 12-MAY-1995		
		APPLICATION NUMBER: US 08/089,372		
		FILING DATE: 08-JUL-1993		
		APPLICATION NUMBER: AU PL4081/92		
		FILING DATE: 14-AUG-1992		
		ATTORNEY/AGENT INFORMATION:		
		NAME: Treccartin, Richard F.		
		REGISTRATION NUMBER: 31,801		
		REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS		
		TELECOMMUNICATION INFORMATION:		
		TELEPHONE: (415) 781-1989		

:		TELEFAX: (415) 398-3249	
:		TELEX: 910 277299	
:	?	INFORMATION FOR SEQ ID NO: 40:	
:	?	SEQUENCE CHARACTERISTICS:	
:	?	LENGTH: 1704 amino acids	
:	?	TYPE: amino acid	
:	?	TOPOLOGY: linear	
:	?	MOLECULE TYPE: protein	
:	?	SEQUENCE DESCRIPTION: SEQ ID NO: 40:	
:	US-08-485-355B-40		
	Query Match	5.7%; Score 515; DB 4; Length 1704;	
	Best Local Similarity	21.7%; Pred. No. 8.8e-36;	
	Matches 397; Conservative 179; Mismatches 553; Indels 700; Gaps 84;		
Oy	13 AAANSALNAVYVR-----PFLSRVOTELLINIMOROLVFREVL--WNHPL----Q 59		
Db	13 AADAVYANLVIOQRVKLDFAPRLKALT-LHRLYTP--LRFKGTLPPHOPIIACHQ 68		
Oy	60 RVIHNELDYCRARAGRCLEGVAHDRS---INDNPVN---LRHCFEIRPGVDYRWYS 112		
Db	69 RVAEVLNIPARGRS-TYLEIGPSLSALKHLGAPNAAYVDHYGC-TYKGRDGSRHITA 126		
Oy	113 PTRGPANCRRSAALGLDP-----ADRTYCEDFSRCFAAETGVALSLLDL 160		
Db	127 LE-----SSSVATGREPFKADASLANGLANGISRFPCVDGYSCAFKRVGIANSLYDV 179		
Oy	161 WPAADVAAEMAHNGXRLYAALHPREVLLPCTYTHTTSYLIIH-----DGD 206		
Db	180 TLEEIANPEFNHGCHMVARAFMHMBPELLYMNVYNAELGYRPHVEEPMAVAKCAFQCGD 239		
Oy	207 -----RAVYTEGDTSG--YNHDVSILRAMIRTTK 235		
Db	240 LRLHEPELFINSOERIERLIANGSYSRAVIFSGDDMDGDAYLDHFHMWLAVL---- 295		
Oy	236 IYGDR-----LVIERRAICGHVILLTAPESPMPYPYPYPRSTEVYRSITFGPGS 289		
Db	236 LVARNPTPEGFSLHIEVGRHHGSSIELKITRAP-----PGDMIAVVPPTSQGLCRI 347		
Oy	290 PSLF--PSASCSTKTFFHAVPVHIMDRMLFGATLDDQAFC-CSRLMTYLRGISYKVTVGA 346		
Db	348 PNIFYAASGTGEHKTIILTSHKHVMMLNMFTQREKELYDMVTIVMSPARAIRAIYVAS 407		
Oy	347 LYANDGNWASDALTAIXITTAAYLTICHORYLRTQAISGMKRLGVEHQKFTRLRYSLMF 406		
Db	408 EYTESMWNISPADLVTRTVASLYLHIIR-----BRAAVA----- 442		
Oy	407 EKSGRDYIPRGLOLPFYAACRRWLMSGFHIDRPVLVFDESVCRCRTFLFKYAGKCCEPMR 466		
Db	443 VKTKADVV-----FGISTFWES-----LKHYLG-SCCCLR 471		
Oy	467 WLQDECTCLEBPAGLVGDHNDNEAVEGSEVDAPEAPHLDVSGTYAVAHQHOLEALYRAL 526		
Db	472 NL-----KGTDV----- 478		
Oy	527 NVFPDIARASRLATATVELVASPDRLBECRYUGNKTFRTYYVDGAHLEANPREOVYLSFD 586		
Db	479 -----VFKRVVADKYV----- 490		
Oy	587 ASROSMGAGSHSLTYELTPAGIQVRISSNGLCDTATPPGGAPSMAAPEVAFCSATLRY 646		
Db	491 -----HSL-----GDI--ICD----- 499		
Oy	647 NREFORMSLNLGLMLHREGGLIGIPPSPSGIHVESANPFCGEGTLYTRTWSTSGSFSDFS 706		
Db	500 -----VRISPFOV-----GF----- 509		
Oy	707 PREAAPAMMATPGJPHSTPVSIDIWVLRPSEEOY-----DAAPVPARPDAQLPBG 759		
Db	510 -----LPKSVPPAR-----VFHDBEELEVLRKAGCYNERPVPST----- 543		
Oy	760 PVVLTPPPPPVHKPSIAPPSPRNRLLYTTDGAKVYVAGSLFEESDCDWLVVANSPGHRPG 819		


```

Db 544 -----PVEBP-----QGDADL-W-----557
Oy 820 GGLCHAFYORPEAFYPTFEMREGIAVATLPPRIHAVA---PDYREONPKRLAAY 876
Db 558 -----HATAASLPEYRA-----TLQAG 574
Oy 877 RETCSRGTAAPLJOSGIYQVPSISPFAMERNRPGDELYLTERAAMFPAKRAQOV 936
Db 575 -----LNTDVKQKLTIL-----586
Oy 937 LTTEDTARTANLAEIDATEVGRACAGTISP--GIYQFTAGVPSGKSRSI-----990
Db 587 -----ENAKT-----ID-----GLTSPVAGLEMYE---GPPSGKGTITIAL 623
Oy 991 -OQGDVVVVPTRRLNSWRRR--GFAATPPHATVATIGRR-----VVIDEA 1036
Db 624 EAAGKATVATPRLREAMDRIKPPSASATQVHA--LAILRATAGAFATVVIDEC 681
Oy 1037 PSLP-PHILLH-MORASSVHLIGDNOIPATIDEHAGL-VPAIRP--ELAPTSMMXVTH 1091
Db 682 FMPPLVVAIVAHLSRSSRIVLGDVHQIGFIDQGSANMPLVRDVVKQCRRTFNQTK 741
Oy 1092 RCPADV--CELIRGAVPKIQTTSRVLSLFMNEP---AIGOKLVYQAANKAN--PGAI 1143
Db 742 RCPADVATTFQSLYPGCTTSGCVASISHVAPDRNSQAOITLCTQBEKSHGAEAM 801
Oy 1144 TVHEAGATFTETTTI--ATADARGLIOSSRAHAIVALTREKCVILDAAG-LIREVG 1199
Db 802 TVHEAGRTFAVILHYNSTAEOKLLAEKS--HLVGTIRNHLIYRIDPGIDIEROL 859
Oy 1200 ISDVIVNPFLLGGEVGNHRP-----SVIPRGNPDONLGTQAPPSOI--SAHQLA 1251
Db 860 HS-----AKAEFTDIPAPLEITTVKPSSEVORN--EVMAITPQSATPHGAILHLR 909
Oy 1252 EELGHRP-APVAALVPPCELEGLYMPQELTVSDV---LVEELTIVHCRMAAPSO 1306
Db 910 KNFGQPDCCGVALAKTGEVFGRAKINVELAEPDAPKPHAFQ--EGVQWVKVINA 968
Oy 1307 KRAVSLTVGRYGRRT--KLYEAHSDVR---ESLAPITPTGVOATTCELYELVEAM 1360
Db 969 KHAQALOTLLSRYSKASADLPHE-AKEDVKRMLNSLDRHMDMTVTEDADRAVETOLKF 1027
Oy 1361 VKKGDGSAVLELDCNRVSRITTFQKCKNCTTGTETIANGVCGISANMSTFECALG 1420
Db 1028 TORGTVEEDLEPD--DPYIRIDIFLMTQOK--VSPKPINTGKVGGLAHSLSLNFVLA 1084
Oy 1421 PWFRAIEKEIILALPPNIFYG---DAYEESVFAAVSGASCMPENFSEFSTQNNF 1476
Db 1085 AMIRILE-ELKGTGRTVYNSGLPDEEAMLEAKINOVPHATFVSADWTEFDTHHNT 1143
Oy 1477 SLGLECVNMECGMPQMLRLYLHVSAMILQAPKESLGFWMKH-----SGEPSTLTM 1530
Db 1144 SELLEPAALLERIGTPAAVNLFRERCGKRTLRA-----KGLGSVEVDGLDSSAAMTPCR 1198
Oy 1531 NNVMMKATIAHCEYERPDFAVAFKGDSDVYLCSDRQSNAAALLAGGGLK---LKVDYR 1587
Db 1199 NTFESAAMVLT--PRGVYFAAFKGDG--LTCGSHYLRDASLHMGERYKKHKLKEVO 1255
Oy 1588 PGLAGVAVVAGLGTLPVVRFA---GR--LSE-----KNNCP-----1621
Db 1256 KIVPTIGLLVSAQOYVL-DPVRSALKIFGRCTSELLISKIYEAVADITKMSDARYHSL 1314
Oy 1622 -----GPERAEQLRALVCOFLRG 1639
Db 1315 LCHMSACYNYVAPESAAYIIDAIVVREGRG 1343

```

RESULT 8
 US-08-093-453B-2
 ; Sequence 2, Application US/08093453B
 ; Patent No. 5439814
 ; GENERAL INFORMATION:

```

APPLICANT: Frey, Terry K.
APPLICANT: Dominguez, Geraldina
APPLICANT: Wang, Chin-Yen
TITLE OF INVENTION: Modified Infectious Rubella Virus
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESS: Jamie L. Greene, Jones & Askew
STREET: 191 Peachtree Street, 37th Floor
CITY: Atlanta
STATE: Georgia
COUNTRY: United States
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50
COMPUTER: Macintosh
OPERATING SYSTEM: 7.0
SOFTWARE: Microsoft Word
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/093,453B
FILING DATE: 19 JUL 1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: U.S. 07/722,334
FILING DATE: 28 JUN 1991
ATTORNEY/AGENT INFORMATION:
NAME: Greene, Jamie L.
REGISTRATION NUMBER: 32,467
REFERENCE/DOCKET NUMBER: 07362-0101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404.818-3700
TELEFAX: 404.818-3799
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2205 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Rubella virus
STRAIN: Thierien
US-08-093-453B-2

Query Match 4.2% Score 374.5; DB 1; Length 2205;
Best Local Similarity 19.2% Pred. No. 3.2e-23;
Matches 454; Conservative 201; Mismatches 714; Indels 995; Gaps 104;

Oy 3 ITTAIDQALAAANSALANAVVVRPLSRVQTEILINLMQPROLVFRPEVLMNHPQRYI 62
Db 40 VYTAAGKRAIV-----AVIPRPVFTQMVS-----DHP---AL 69
Oy 63 HNELEQYCARAGRCLEVGANPRS-----INDPNVLHRCLELPVGRDYQKWS-----111
Db 70 H-AISYTRR--HWTEWG--PREALHVLIDPSGL-----LREVARVERRWALCLIHRT 118
Oy 112 -----APTRGPAN-----CRSALRGLP-----ADRTYC-FDGF 141
Db 119 AKRLATALETSEAMHADIYC---ALRGAPSGPFYVHPEDVPHGGRAYVADRCLLYTTPM 175
Oy 142 SRCFAETGVALYSLHDEMPADVA-----E 167
Db 176 QNCELMRTIDATLTLVAVDMLPVALAHHVGDMDMDLGIAMHLDHDSGCPADCRGAGGPTP 235
Oy 168 AMARHCXTRLYAALHLPREVLLPFGTYHTTSTYLLIHDSRAVVTYEGDTSAGYNHDVSLI 227
Db 236 GYTRPCYTRIIYOVL---PDTAHGRLYRCGPRMLTWDCAVAELSWVAOHCQHQ-----286
Oy 228 RAMIRTKIVGDHPLVTEVRAIGCHFVLLTLTAAPSPMPVYVPYRSTRSEVYVRSIFGPG 287

```

287 -----ARVAVRCLPIRHVRSLOPS-----ARVR-----311
288 GSPSLFASCKSTFHAVVHMDRLMFGATLDQAFCCSRMTYLRGISYKVTGAL 347
312 -LIDLVLAEVGRKRWISLRPFVQRMISYCKTISPRAYSEREFKKNALCHSITLACN 370
348 VANEWN---ASEDALTAITTAAYLTTCORLYLTQAIISGMRLGVEHAKFTRLYS- 403
371 VLOGMGTCAEEDALCA-----YVAFRAMQSNARLAGIMKAKCAADSLAV 417
404 ---MLFEKSGRDYIPGRLOFYAOCRRMLSA-----GFHLDRVLYFESVRCRTE 453
418 AGML-----DTI-----WDAIKRFLGSVPLAERMEBQDAVAADF-----454
454 LKRVAGKCCFMRWLGEOCTCFLEPAGLVGDHGDNEAVEGSEVDAEPAHLDVSGTYA 513
455 -----RGPLEDGGRHLDYQ-----PKSPRPRIATW 484
514 VHGHOLE---ALYRALVPOD-----IAARAS 537
485 VHAASEDRHCACAPRCVPRERPSAPAGPDDEALIPMILFAERRALCREMFEALRAR 544
538 RLATFVELVASDRLBECRTVL-----GNKTFRTYVDGAHLEANGPEOY 581
545 ADTAAAP 604
582 VLSEFASROSMGAGSHSLTYELTPAGLO--VRISNGLDCTATFPFGAPSAAPGEVAAF 639
605 -----FAAGAHMCA---QARGLOAFVRY-----VPRPERWADGG--ARA 639
640 CSALYRNRTQORHSLTGMLHPEGILG-----IFPPSGHWTESANPCGECTLYRT 695
640 WAFEFRCACAAOR-----LLGEPYVHMLTYTGDVPOQL-----ALALRT 679
696 WSTSGFSDS-----PPEAAPAMAAT-----PGLP-----HS 724
680 LAOGGALALSVNDLPFGAIFDANAAYTAARAGROSAASPPGDDPPRRARSRHS 739
725 ---TPVSDIWLPPSESEFOVADAPVPAAPDAGLPGVVLTPPPPVHK-----773
740 DARCTPPAP 787
774 ---PSIPEPSRNR-----RLLYTPDGAKYKASLESFSDCWL 808
788 VACEPSGPTSTRADPDSIDYESTARAGVHLARVDIMDPPCKV-----V 835
809 VNASNPGRGGGLCHAFYORFPEAFYTFEIMREGLA--YTLTPRP-----854
836 VMANEGLLAGSGVCAIFANATA-----LAANCRLAPCPTEBAVATPGHGC 884
855 ---IHAVAP-----DYVEONPKLEAAYRE-----TCSRKGTAAVPLIGSIT--QV 898
885 GYTHIHAVAPRRBRDPALAEGBALLERAYRSIVALAARRMVACVCPILGAGVYWSA 944
899 PVSLSFDAMERNRPGODELT-----TEPAAWFE 928
945 AEBLRAALAAATREVERVSLHICHPDRATLTHASVLYGAGLAARVSPPTLELAS-CP 1003
929 ANKPAOPVLITTEDTATAMALEIDATEVGRACAGTIS-----969
1004 AGBPGRAGQASAPPAITPLG---DATAPEPCQGCCLCRTLRYTNDAYVNLMLERD 1058
970 -----PGIV-----HYOF-----977
1059 RGATSNAMRIPEVYVYVYBEHLATHFPLNHVSVLKPAEVRPPRCMGSGDMRCRGMHPQ 1118
978 -----TAGVPGSGKSRSTOQGDVVDVYVPTRELIN--SMRRGFAAFTPTTAA 1023
1119 VRCPTPSNAHALCRTCQVPPASTRG---GELDPNTCWLRAAANAQAARACGAYTSAGCP 1175
1024 RVITGR-----RVVIDEAP 1037
1176 KCAVGRALSEARTHEDEFAALSQRMSASHADSPDGTGDPIDPLMETYVGCACSRVWVGSSEH 1235

1038 SLPPHLLHLMORA-----SVVLL-----GDPNOI---PAI-----1066
1236 EAPPDHLLVSLHRAPNRPMGVLEVRRARPEGNPTGFCVAVGSGPRRVSDBRHMLAVP 1295
1067 -----DPE-----HAGVLAIRPELAIPS--WXYT---1090
1296 LSRGGTCAATDEGLAOAYVDLEVRRLDGDAMARALASVORPRKGPYNI RWNMAAGA 1355
1091 -----HRCPA-----DVCELR-----GAYPTI--1108
1356 GKTRILIAATFREDLYCPCPNALLHEIQAKLRARDIDIKNAATYERLTPLAAYRIYI 1415
1109 -----QTSRVL-----RSLF--NNEP--1123
1416 DEAFITLGEYCAVASQTTAEVICVGDROCGPHYANNCRTPVDPDRWPTBRSHRTWFPD 1475
1124 -----AIGKLYXTQAARANKAIGATVHE 1147
1476 CMAARLRLAGLDYDIEGERTGFACNMDRGVDLHLAFSRETV--RRLHBAGIRAYTVRE 1533
1148 AOGATFETETIATADARGL-IOSRAHALVALTRHTEKCVIID-APGLREVISDVIY 1205
1534 AOGMSVGTACIHYGRDGTVALALTRDLALVSLFRASDALYHLEDEGSLRAAGLS-----1589
1206 NNEFLAGEV-----GXNR-----PSVIRGNPDONICTLOAFPSCQISAYHOL 1250
1590 --AFLDGALAELEKEVPAGIDRVAVVQAPPLP---PADGIEADQVPPFCRTLEELV 1644
1251 AEBLGRPAVAVALPPCPLEBOGLYMPBELVSDSVLYFELTDVHCMAAPSORKAY 1310
1645 FGRAGH---PHYADLNKVTGEREVRVM---RISRLHLKNHTEM-----PGTER-V 1689
1311 LSTLVGRYGRRTKLYEAAH--SDVRESLARFIP---TIGPVQAT---CELVEL 1356
1690 LSAVCA-----VRRYRAGEDSGLTRTAVARQHPRFPFQOIPPRVYTAGVAGEMHTYLRER 1744
1357 VEAMVEKGQDSAVLEL-----DLGNRVSRTTFQOKCNK---T 1394
1745 IDLTDVYTMGVARELTDYRARRPEIFAGMCTAOSLSVAPAFKATLKCVDAALGPRDT 1804
1395 TGETIHAKGVGOGISAMSKFCALFGFWFAIREKELIALLPNIIFGDAAVEESVFNAAYS 1454
1805 EDCMAOGKAGLEIRAMAKMVQVMSHPRAIOKILMRALRPOFLVAAGITEVEVDAMWQ 1864
1455 GAGSCWVENDFSEPDSTONFS-----LGLECYVMECGPOMLIRLHYLVRS 1503
1865 AHYTTNAIEVDFTPEFDMNQTLATRDVELISAALLGPCA--ED-----YRALRA 1912
1504 -ANILDAPKESLKGFMKKHSGEPGTILMNTVMMALIAHCYER---DPRVAA--FRGDD 1557
1913 GSYCTLRELGSTETGERTSGEPATLLHNT---VAMCMAMRMPKGVRMAGITQGD 1967
1558 SVVLCSDYROSNAAL-----IAGCKLKLVDYRPIGLVAGVYVARGLSTLPDVR 1609
1968 MVFLPEGARS---ALKTKPAEYGLFGFHIPYKHNSTPSPSCG--HYGTAAGLFDVYM 2023
1610 FAGRLSEKMGPPPERAEOURLAVCDFLRLGTLNVAQVCDVVSRYGVSPGLVHNLIGML 1669
2024 QAIKVLCCRFE--DPDVLDEEQVALLDRLRG-----VYAAALP-----2057
1670 QTTADGKAHFTETIKPYLDLTNSI 1693
2058 DTVANANAAYDYSAERYLAIVREL 2081

RESULT 9
US-07-876-941A-31
; Sequence 31, Application US/07876941A
; Patent No. 5885768
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R.
; APPLICANT: Bradley, Daniel W.

Db 1709 TDISLMDSESGSLFSSFGSDNYRQVYVADVHAOEPAPVPPRLKMARLAARMQ 1768
QY 1081 LAPTSMXVTHRCPADYCELLRCAYPKIOTSRVLSLFNNEPAIGOKLVYTOAKANP 1140
Db 1769 EEPFP-----PASTSSADESLHLSFDGVSISFGSLFDGEMA---RLAAQPPASTCP 1817
QY 1141 -----GATVHE---AGATFETETIIAT---ADARGLIOSRAHALV-----AL 1179
Db 1818 TDVPMFSGSFSDGEIELSRVTESEPVLFSGFEPGEVNSIISRSVSPPRKORRRR 1877
QY 1180 TRHTEKCV-----ILDAPGLREVGISDVIYNNFPLAGGEVGHRSVITRGNDQ 1230
Db 1878 SRREYCLTGVGGIYFTSDGPGHLOKKSIVLQNOLEPTLERVLEIYAVLDTSKEEQ 1937
QY 1231 NLGTLQAFPPSCQISAVHOLAEE-----LGHRAVVAVALPPC-----PELE 1272
Db 1938 LKLRQAMPPEANKSRQSRKVENQKAITTERLLSGRLVNSAIDQECYITTPKRSYS 1997
QY 1273 QGL--LYMPELTVS-----DSVLVELTD-----IVHC-----1299
Db 1998 SSVAPANSDFPAVAVCNMYLHENYPTVASYQIDEDDAYLDMDVGTACUDTATFCPAK 2057
QY 1300 -----RMAAPSORKAVISTLVGRGRKLYEAHSDVRESLARIPITIG-----1344
Db 2058 LRSTPKRHEHRAPIRIRSAVPSAMONTL--QNVLIATIRKNCNTQIMELPTLDSATNVE 2115
QY 1345 -----PVQATTCELYELVEMAVE--KGODGSVLE-----LDLCNRD 1379
Db 2116 CFRKYACNDEWEFEARKPIRITT---EFVYAVVARLKGPKAALFAKTHNLVPLQEV 2171
QY 1380 VSRITFEOKKCNKFTTG--ETIAHGKVGQGSAMSKTFPCALFGWFRATIEIITALLP 1437
Db 2172 MDRFVMDKRDVKTPTGKTHHEERKV-QVYQAAEPLATATLGLIHVELVRLTAVLLPN 2230
QY 1438 I---FYGDAVEESVFAAASGAGSCWFEENDESEFSTONNFSGLCEVWEECGMPQWL 1494
Db 2231 IHTLFOMSAEDFDALIHNFHQSD-PYLETDIASFDKSQODAMALTGIMLELDGVQOPL 2299
QY 1495 IRL---YHLVRSAMITQAPRESLKGFWKKGSGEPGLTMMTVNMAIIAHCYE--FRDF 1548
Db 2290 LDLECAFGELISSTYHLPDGTGRFKFGAMMK--SGMFLTELVTVLVNVIASVLEERLKT 2347
QY 1549 RVAFKDDDSV--LCSDYRORSMAALLIAGCGKTK-----VYRPLGLYAGVVAPGL 1601
Db 2348 KCAFAIDDDNLIHGVSVDKEMAEKATWL---NMEVKIIDAVIDGERPPYFCGGIILDSV 2404
QY 1602 GLTLDVVRAGRLSE--KMGPCP---ERAEQLRLAYCD-----FLGLFTVAVQVDV 1651
Db 2405 TS--TACRVADPLKRLKRLGKPLPADDEQEDERRALLDETKANFRVGITDITLAVANATR 2462
QY 1652 SRVYVSPGLVHNLIGMLQTIADGRAHF 1679
Db 2463 YEVDNITRPVLL---ALRFEAQSKRAF 2485

RESULT 11
US-09-102-248-2
Sequence 2, Application US/09102248
Patent No. 6008035
GENERAL INFORMATION:
APPLICANT: Johnston, Robert E.
APPLICANT: Davis, Nancy L.
APPLICANT: Simpson, Dennis A.
TITLE OF INVENTION: System for the In Vivo Delivery and
TITLE OF INVENTION: Expression of Heterologous Genes in the Bone Marrow
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bell Seltzer Park & Gibson, P.A.
STREET: 1211 East Morehead Street
CITY: Charlotte
STATE: No. 6008035th Carolina
COUNTRY: USA
ZIP: 28234

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/102,248
FILING DATE:
CLASSIFICATION:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/801,263
FILING DATE: 19-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sidley, Kenneth D.
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5470-147
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175
INFORMATION FOR SEO ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2500 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-102-248-2

Query Match 2.64; Score 233; DB 3; Length 2500;
Best Local Similarity 20.44; Pred. No. 1,1e-10;
Matches 238; Conservative 141; Mismatches 457; Indels 332; Gaps 51;

QY 803 SDC--DWLVANSNGHRRGGGLCHAFYQREPEAFYPTFEITREGLAAYTLP--RPIIHAV 859
Db 1359 ADCGEAVVNAANPLGRGEGVCRAIYKRWPNSE--TDSATETGAKLVCOGKKVIAV 1416
QY 860 APDVRVQNP-----KLEAYRETCS-----RGTAAVPLIGSGLY-----QVPSLS- 903
Db 1417 GPDR--KHPEALKILLQNNAYHVAADVNEHNIKVAIPLLTSTGIYAGKDRLEVS LNC 1474
QY 904 -----FDA-----W-----ERNHRGDELVTPEPA-----924
Db 1475 LTTALDRTDADVTYICLDKKKKEKIDAVLQKESVTELKDEMDIIDLWVIMHDSCLKG 1534
QY 925 -----NMFEANKPAQPVLTTEEDTARTAN-----LALIEDAATEVG 960
Db 1535 RKGSTTKGLKLYSEFGTKFHQAADKMAELKVLFPNDQESNEQCAIILEGTMAINE-- 1592
QY 961 RACAGCTIS-----PGIVHYOFTAGVPSGSKRSIQGDV--DVYVVPTRRLNS 1008
Db 1593 ---KCPVDHNPSSSPPTLPLCLCMYAMTPERVRLRSNNVKEVTYCSSTPLPKYIKNV 1648
QY 1009 WRRR--GEAFTPTTARAVTIGRRVVIDEAPSLPHL-----1043
Db 1649 QKVOCTKVVLFPHTPAFVPAKTYTEAPQAPPAQAEAPGVAVTPTPPADNTSLDV 1708
QY 1044 ---LLHMORAS-----SVHLGDPNPQIPADIFENAGLVAIRPE 1080
Db 1709 TDISLMDSESGSLFSSFGSDNYRQVYVADVHAOEPAPVPPRLKMARLAARMQ 1768
QY 1081 LAPTSMXVTHRCPADYCELLRCAYPKIOTSRVLSLFNNEPAIGOKLVYTOAKANP 1140
Db 1769 EEPFP-----PASTSSADESLHLSFDGVSISFGSLFDGEMA---RLAAQPPASTCP 1817
QY 1141 -----GATVHE---AGATFETETIIAT---ADARGLIOSRAHALV-----AL 1179
Db 1818 TDVPMFSGSFSDGEIELSRVTESEPVLFSGFEPGEVNSIISRSVSPPRKORRRR 1877
QY 1180 TRHTEKCV-----ILDAPGLREVGISDVIYNNFPLAGGEVGHRSVITRGNDQ 1230
Db 1878 SRREYCLTGVGGIYFTSDGPGHLOKKSIVLQNOLEPTLERVLEIYAVLDTSKEEQ 1937
QY 1231 NLGTLQAFPPSCQISAVHOLAEE-----LGHRAVVAVALPPC-----PELE 1272

Db 2394 IIDAVIGERPPYFCGGFILLQDSVTS--TACRVADPLKRLFKLGRPLPADDEDDEDRRL 2451

QY 1634 CD-----FLRGITNAQVCVIVSVRYGVSPLVNLIGMLQTIADGKAHF 1679

Db 2452 IDETRAMFRVGTITGLAVALVATREVDNITPVLL-----ALRTFAOSKRAF 2497

RESULT 14

US-08-801-263A-5

; Sequence 5, Application US/08801263A

; Patent No. 5811407

; GENERAL INFORMATION:

; APPLICANT: Johnston, Robert E.

; APPLICANT: Davis, Nancy L.

; APPLICANT: Simpson, Dennis A.

; TITLE OF INVENTION: System for the In Vivo Delivery and

; NUMBER OF SEQUENCES: 12 Expression of Heterologous Genes in the Bone Marrow

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Bell Seltzer Park & Gibson, P.A.

; STREET: 1211 East Morehead Street

; CITY: Charlotte

; STATE: No. 5811407th Carolina

; COUNTRY: USA

; ZIP: 28234

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent in Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/801,263A

; FILING DATE: 19-FEB-1997

; CLASSIFICATION: 514

; ATTORNEY/AGENT INFORMATION:

; NAME: Sibley, Kenneth D.

; REGISTRATION NUMBER: 31,665

; REFERENCE/DOCKET NUMBER: 5470-147

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919-420-2200

; TELEFAX: 919-881-3175

; INFORMATION FOR SEQ ID NO: 5:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2517 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-801-263A-5

Query Match 2.4%; Score 212.5; DB 2; Length 2517;

Best Local Similarity 19.7%; Pred. No. 7.3e-09;

Matches 235; Conservative 141; Mismatches 454; Indels 361; Gaps 50;

QY 803 SDC--DWLVNANPGRHPRGGLCHAFYOFEPFAFYPTFEIRMEGLAAYLTLP--RPIIHAY 859

Db 1359 ADCQEAHVYVNAANPLGRBEGCRALYKRWPNF--TDSATETGTRAKLTVCQCGKAYIHAV 1416

QY 860 ADDYVEQNP-----KRLAAVRETCS---RGTAAVPLLGSGTY-----QVPYSLS- 903

Db 1417 GPDFR--KHPEALAKLLQNAVHAVALVNEHNKISVAIPLLSTGIYAAGKRLREVSILNC 1474

QY 904 -----TDA-----W-----ERHNRGDELYLEPPAA----- 924

Db 1475 LTTALDRDADADVTYCLQKKMERIDAVLQIKESVIELKDEDEIDDELWTHIPDSCLKG 1534

QY 925 -----NMFEANKPAQPVLTTFEDTARTAN-----LALRIDATEATWG 960

Db 1535 RKGFSSTGKGLKLYFEGTKFHOAKDMAELIKLFPDDESINOLCAVILIGETMEAIRF-- 1592

QY 961 RACAGCTIS-----PGIVHQTAVGVGSGKSRHSIQGGV--DVVVVPTRELANS 1008

Db 1593 ----KCPVDHNPSSSPKTLPLCLCMVAMTPREVRHRLRSNNVKVTCSSSTPLPKYKIKNV 1648

QY 1009 WRR--GFAATPHTAARTYIGRRVVIDEAPSLPHL----- 1043

Db 1649 QKVCCTKVVLFNPHPAFVAPARKYIEAPEQAPAPQAQEADEAVATPPPADNTSLDV 1708

QY 1044 --LLHMPAS-----SYHLGDPMQ 1062

Db 1709 TDISIDMEDSSRGSLSFSSSTGSDNITSMDSSSGPSLSLEYDRQVYVADVAHVAQEPAP 1768

QY 1063 IPADIFEHAGLVPAIRPELAPTSMWXVTHRCPADVCEILRGAYPKIQTTSRVLSLFWNE 1122

Db 1769 VPPRLKMMARLAARMQEEPFP-----PASTSSADESLHLSFGVSGMSFGSLFDE 1820

QY 1123 PAIGKIVYTOAKKANPAIT-----VHE-AQCATFETITITADAR 1165

Db 1821 WG-----ALAAQAPASTCPIDVPMSFGSFGDEIEELSRYVTESEPVLFSGFERG 1871

QY 1166 GLIOSSRAHAIVAL-----TRHTKCYIL-----DAPQLREVGISDIYVNN 1207

Db 1872 EVNSIISKSVSFPFRKQRRRRRRRTYELTGVCYIISTDTGRHLOMESVLQNLTE 1931

QY 1208 FFLAGGEGXHRPSYIPRGNDPDLGTLQAPPPSCQISAYHOLAE-----LGH 1256

Db 1932 PTLERNVLEIRIYAPVLDTSKEQOLKLRVQMPTEANKSRQSRKVENOKAITTERLLSL 1991

QY 1257 RAPVAVALPPC-----PELEQGL--LYMPOELTVS-----DSVLYFELTD-- 1295

Db 1992 RLVSATDQPECKYKITYPKPSYSSSVANYSDBPKRAVACNNYLHENYPTVASYOITDXY 2051

QY 1296 -----IVHC-----RMAAPSORKAVLSTLVGRYGRRTKYEA 1328

Db 2052 DAYLQVDSVACLOTATTCPRAKLSYPRKHREYRAPNTSAVPSAMQNTL--QNTVLAAT 2109

QY 1329 HSDVRESLARFIPTIG-----PVQATTCLEYELVEAMVE- 1362

Db 2110 KNCNVQTMRELPTLDSATFNVECFKRYACNDEYWEFAKPKDIRTT-----EFTAYVAR 2165

QY 1363 -KGQGSAAVLE-----LDCNRDVSRTTFQKXCKNFTTG--ETIANGVGGISMSMT 1414

Db 2166 LKGPRAALFAKTHNLVPLQEVPMDFVMDKRDVAVTPTGTHTEERPVP--OVLQAEPL 2224

QY 1415 PCALGPFMEFRAIEKEILLALPPNI---FYGDAYEESVFAAASGASGVFENDESFPDS 1471

Db 2225 ATAVLGCIHRELVRRLTAVLLNINHTLFMSAEDFPAIIAHRKQGD--PVLEIDIASFPK 2283

QY 1472 TQNNFSLGLECVAMEECGPMQLIRL---YHIVSAMLLQAPKESLKFWMKKHSEPGT 1527

Db 2284 SODDAMALTGMLIEDLDGVDOPLDLIECAFGEISSTHLPGTGRFKFGAMMK--SGMPLT 2341

QY 1528 LLMNTVWMAAIIAHCE--FRDFRVAAPFKGDSVY--LCSDRQSNMAAALIAAGGLKIX 1583

Db 2342 LVTNVTLVNVIASRYLEERLKTSCAFTGDNITHGVVSDKEMARCATWL---NMEVK 2398

QY 1584 ----VDYRPIGLYAGVVAAPGLGTLPDVVRPAGRLSE--KMWGPP---ERAQDLRLAV 1633

Db 2399 IIDAVIGERPPYFCGGFILLQDSVTS--TACRVADPLKRLFKLGRPLPADDEDDEDRRL 2456

QY 1634 CD-----FLRGITNAQVCVIVSVRYGVSPLVNLIGMLQTIADGKAHF 1679

Db 2457 IDETRAMFRVGTITDLAVALVATREVDNITPVLL-----ALRTFAOSKRAF 2502

RESULT 15

US-09-102-248-5

; Sequence 5, Application US/09102248

; Patent No. 6008035

; GENERAL INFORMATION:

; APPLICANT: Johnston, Robert E.

; APPLICANT: Davis, Nancy L.

; APPLICANT: Simpson, Dennis A.

; TITLE OF INVENTION: System for the In Vivo Delivery and

; NUMBER OF SEQUENCES: 12 Expression of Heterologous Genes in the Bone Marrow

Thu May 31 09:15:06 2001

us-09-468-147-91.raii

Page 16

```

CORRESPONDENCE ADDRESS:
ADDRESS: Ball Seltzer Park & Gibson, P.A.
STREET: 1211 East Morehead Street
CITY: Charlotte
STATE: No. 6008035th Carolina
COUNTRY: USA
ZIP: 28234

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/102,248
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/801,263
FILING DATE: 19-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: Sibley, Kenneth D.
REGISTRATION NUMBER: 31,665
REFERENCE/DOCKET NUMBER: 5470-147
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-420-2200
TELEFAX: 919-881-3175

INFORMATION FOR SEO ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2517 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-102-248-5

```

	Query Match	2.4%	Score 212.5:	DB 3:	Length 2517;	
	Best Local Similarity	19.7%	Pred. No. 7,3e-09;			
	Matches 235;	Conservative 141;	Mismatches 45;	Indels 361;	Gaps 50;	
QY	803 SDC--DWLVNNSNPCHRRGGGLCHAIFYORFPEAFYPTETIMEEGLAAYTLRP-RPIIHAV	859				
Dd	1359 ADCCOEAAVVAANPILGRPEGVCRAIYKRKPNSSF--TDSATETGTAKTLVCGCKKHIVAH	1416				
QY	860 APDYVEQNP-----KRLAARFECSS-----RRTGAAPLLGSGIY---QVPYSLS-	903				
Dd	1417 GPDEF--KHPEALKLIONAHAVADLVNEHNISVALPILSTGISIYAAGKDRLSEVSLNC	1474				
QY	904 -----FDA-----W-----ERNHRBODELYLTPAA----	924				
Dd	1475 LTTALDRIDADVTIYCLDKKKMERIDAVLQLKESYIELTKDEMEIDELVMWHHPDCLKG	1534				
QY	925 -----NWFEAKKPAPQVLITTEDPARTAN-----LAEITDATEYG	960				
Dd	1535 RKGFSTTGKGLYSYEGTKFHQAARDMAEIKVLPENDQESNBOLCAYLIGETMIRE--	1592				
QY	961 RACAGCTIS-----PGIVHYOFTAGVGSGSKRSIOGDV-DVVVVPTRELNRS	1008				
Dd	1593 ----KCPVDHNPSSSPKTLPLCLCMYANTPREVNHLRSNKNEYVCSTPLPKIKINV	1648				
QY	1009 WRRR--GFAAFTPHTAARVYTIGRRVIDEAPSRLPHL-----	1043				
Dd	1649 OKVOCTKVLVLFPNHPTRAFVPAVKYTEAEOPARPAPAOAEAEVAATPTPPAADNTSLDV	1708				
QY	1044 ----LLHHMRAS-----SVHLGDPHQ	1062				
Dd	1709 TDISLMDDESSEGLSFSSFGSDNSTMTSDMSGSPSSLLETIVROVVADVHAHOEPER	1768				
QY	1063 IPAIDEFHAGLVPAIRPRLAPTSMWXVYNHRCPADVCCELLRGAYPKIQTSRYLRSLFNWE	1122				
Dd	1769 VPPRLKTMARLAARAQOEPT----PASTSSADESHLETFGGVSMSFGSLFDGE	1820				
QY	1123 PAIGOKLVXTHDAKANNGAIT-----VHE-AQGATTEETTIIATPADR	1165				

D	b	1821	MG-----ALAAQPPASTCPTDVPMSFGSFDGBEELBSRATSEEPVLFSFBERG	1871
Q	y	1166	GLISSRAHAIVAL-----TRHTEKCVL-----DAEGLENGISDIYVNN	1207
D	b	1872	EVSNIISRVSAPPKPKORRRRRRTXETLGVGYIFSTDTGRBLOMESVLONOLET	1931
Q	y	1208	FELAGGEVGHBPVIRGNBDONLGTLAFPSCQISAYHQLAEE-----LGH	1256
D	b	1932	PFLERNVLERIYAPVLDTSKEBQLKLRYOMPLPEANKSKRYSRKVENOKAITTELLSL	1991
Q	y	1257	RPAVAVALPPC-----PELBGL- LYMPDELTVS-----DSYLVELJTD-	1295
D	b	1992	RLYSATDOPECXYITFKPXSYSVVANSDPKFAVAVANNYLNHENYPTVASQYTDEY	2051
Q	y	1296	-----LYHC-----RMAASOKRAYLSLILYRGYGRRTLZYEA	1328
D	b	2052	DAYLDWVGTVACLDTPATFCPRAKLRSTPKRHEXRADNTRSAPVSAQNLT--QNVLIAAT	2109
Q	y	1329	HSDRESLARFIPTIG-----PVQATTCELYELVAMEY-	1362
D	b	2110	KRCNCNTQMHELPLDSATFNVECFPKYACNDWEYEFARKPIRLITT-----EFTAYVAR	2165
Q	y	1363	-KGDDGSAYLE-----LDLCNRBYSRITTFQKKCKNFTHG--ETIAHGKYGGGISAMSKT	1414
D	b	2166	LKGRKAALRAKTNHLNLPLOEVMBDRFRVDMKDADVVTGTGHTHERPKV-OVLOAAEPL	2224
Q	y	1415	FCALFFGWMPRAIEKEILLPNI---FYGDAAEESVFMAAASGASCWFENDESFEPS	1471
D	b	2225	AFAVLGCIHHELVRRLTAVALLPINHIHLPMSADEPALIIAEHFPGOD-PYLETDIASDK	2283
Q	y	1472	TONNESLGLEBVMEBCGMQWLIRL---YHLVRSAMILOAPKBSLGFMKHSGBEGT	1527
D	b	2284	SODDAMAALTGMILEDEDGVODPLDLIECAFGEISSITHPTGTRFKFGAMM--SGMFJT	2341
Q	y	1528	LMMTWNMMAIIAHCEY--FRDERVAAFPGDDSV--LCSDVROSRNMAALTAGCGELTK	1583
D	b	2342	LFWNTLVANVIASVLEERKTSCAFATIGDDIINGVSDXEMABRCATWL--NMEX	2388
Q	y	1584	-----VDYRIGLAGYGVYAVALGTLPDYVRFAGRUSE--KNMGGP--ERAEOURLAV	1633
D	b	2399	IIDAVIERPPPYFCGFILODSVTS--TACRAVDPLKRLFKLGKPLPADBEDOEDERRRAL	2456
Q	y	1634	CD-----FLAGLITNAVOVCYDVVSRKYGVSPGLVNHILSMLOTIADGRANF	1679
D	b	2457	LDETKANFRGIDTDLAVAAATRYEVDANTIPVLL-----ALTFPQOSKRAF	2502

Search completed: May 30, 2001, 16:09:45
Job time: 202 sec

XX XX
PN W09919732-A1.
XX
PD 22-APR-1999.
XX
PE 15-OCT-1998; 98MO-US21941.
XX
PR 15-OCT-1997; 9705-0061199.
XX
PA (ABBO) ABBOTT LAB.
XX
PI Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;
XX WPI; 1999-288017/24.
DR N-PSDB; 200195.
XX
PT Detection of United States Isolates of hepatitis E virus
XX
PS Claim 6; Page 175-180; 260pp; English.
XX
CC The invention provides a method for detecting a US (sub)type hepatitis E
CC virus (US-HEV), or its naturally occurring variants in a sample by
CC treatment with a binding partner specific for a marker of the virus, and
CC then detecting any complex formed. The method is used to diagnose
CC infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
CC open reading frames (ORF) in US-HEV and host cells expressing these ORFs
CC are useful in vaccines or for passive immunisation. The polypeptides are
CC also used to raise specific antibodies (useful as immunoassay reagents).
CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
CC usual hybridisation and amplification assays for detecting infection. The
CC present sequence represents a HEV-US1 ORF1 protein.
XX
SQ Sequence 1698 AA:

Query Match 99.9%; Score 9001; DB 20; Length 1698;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PGIITTAEQALAAANSALANAVVRRPFLSRVQTEILINLMQPOLFRREVLMNHP10R 60
DB 1 PGIITTAEQALAAANSALANAVVRRPFLSRVQTEILINLMQPOLFRREVLMNHP10R 60
QY 61 VINHELEQYCRARARAGCLEVGAHPRISINDNPVLIHRCFLRPVGDVORWTSAPTRGPAAN 120
DB 61 VINHELEQYCRARARAGCLEVGAHPRISINDNPVLIHRCFLRPVGDVORWTSAPTRGPAAN 120
QY 121 CRISALRGLPPADRTYCFDFGSRCAFPAETGVVALYSLHDLMPPADVAAMARHGCTRLYYAA 180
DB 121 CRISALRGLPPADRTYCFDFGSRCAFPAETGVVALYSLHDLMPPADVAAMARHGCTRLYYAA 180
QY 181 LHPPEVLLPPTGYVHTSYLLIHDGDAVVYEGDTSAGYNHDSILRAMIRTKIYGDH 240
DB 181 LHPPEVLLPPTGYVHTSYLLIHDGDAVVYEGDTSAGYNHDSILRAMIRTKIYGDH 240
QY 241 PLYIERRAIGCHFVLLTLAAPERSPMPYVYPRSTVYVRSIFPGSGSLEPSSACSTK 300
DB 241 PLYIERRAIGCHFVLLTLAAPERSPMPYVYPRSTVYVRSIFPGSGSLEPSSACSTK 300
QY 301 STEHAVVNHIMDLMLRGATLDDOAFCCSRMLMYLRGISYKVMYGALVANEGMNASDAL 360
DB 301 STEHAVVNHIMDLMLRGATLDDOAFCCSRMLMYLRGISYKVMYGALVANEGMNASDAL 360
QY 361 TAXITAAVLTICHQRYLRTQAISKGRRLGVENAOKITFLYSLWFEKSGRDYIPGRLO 420
DB 361 TAXITAAVLTICHQRYLRTQAISKGRRLGVENAOKITFLYSLWFEKSGRDYIPGRLO 420
QY 421 FYACQCRMLASGHLDBRVLYFDBSVPCRCRTPLKKAAGFCFCCPMRLGOECCTCLEPAE 480
DB 421 FYACQCRMLASGHLDBRVLYFDBSVPCRCRTPLKKAAGFCFCCPMRLGOECCTCLEPAE 480
QY 481 GLVGDGHQNEAEGSVDAEPANLIDVSGTYAVHGHQLEALYRALNVPDDIARASRLT 540
DB 481 GLVGDGHQNEAEGSVDAEPANLIDVSGTYAVHGHQLEALYRALNVPDDIARASRLT 540

DB 481 GLVGDGHQNEAEGSVDAEPANLIDVSGTYAVHGHQLEALYRALNVPDDIARASRLT 540
QY 541 ATVELVASPDRLCECRVTLGKTKTFTTVVDGAHLEANGPEQYVLSFDSRQSMGASHLT 600
DB 541 ATVELVASPDRLCECRVTLGKTKTFTTVVDGAHLEANGPEQYVLSFDSRQSMGASHLT 600
QY 601 YELTPAGLOVRISNSGLDCATTEPPPGAPSAAPGEVNAFCSALYRYNRFTQRHSLTGLW 660
DB 601 YELTPAGLOVRISNSGLDCATTEPPPGAPSAAPGEVNAFCSALYRYNRFTQRHSLTGLW 660
QY 661 LHPBGLLGIFFPPSPRGHWSANPFCGEGTLYRTWSTSGFSDSPPEAAARMAATPG 720
DB 661 LHPBGLLGIFFPPSPRGHWSANPFCGEGTLYRTWSTSGFSDSPPEAAARMAATPG 720
QY 721 LPHSTPEVSDIWLPPPESEFOVDAAPVAPAPAGLPGCVLTPPPPPVHPSPISPPS 780
DB 721 LPHSTPEVSDIWLPPPESEFOVDAAPVAPAPAGLPGCVLTPPPPPVHPSPISPPS 780
QY 781 RNRRLVTPYDGAKKVYAGSLFESDCMLVNASNPNRPPGGGLCHAFYORPEAFYPTFT 840
DB 781 RNRRLVTPYDGAKKVYAGSLFESDCMLVNASNPNRPPGGGLCHAFYORPEAFYPTFT 840
QY 841 MRGGLAATITPRPIIHAAPVDRVEQNPRLAARFETCSRRGTAYPLIGSITQYVPY 900
DB 841 MRGGLAATITPRPIIHAAPVDRVEQNPRLAARFETCSRRGTAYPLIGSITQYVPY 900
QY 901 SLSEFADERNRHRGDELTYLTPPAANPEANKPAOPVLTITEDTARTANLLEIDAATEVG 960
DB 901 SLSEFADERNRHRGDELTYLTPPAANPEANKPAOPVLTITEDTARTANLLEIDAATEVG 960
QY 961 RACAGCTISPGIYHXYOFTYAGVPSGSKRS10QGDVDVVPVTRELNRSWRRRGFAAFTPH 1020
DB 961 RACAGCTISPGIYHXYOFTYAGVPSGSKRS10QGDVDVVPVTRELNRSWRRRGFAAFTPH 1020
QY 1021 TAAARVTIGRRRVIDEAPSLPRLHLLHMORASSVHLLGDNOIPALDFEAGLVPAIRPE 1080
DB 1021 TAAARVTIGRRRVIDEAPSLPRLHLLHMORASSVHLLGDNOIPALDFEAGLVPAIRPE 1080
QY 1081 LAFTSWMXVTHRCPADVCIELRGAPRKIQTSRVLSLFFNNEPAIGOKLXYTQAAKAANP 1140
DB 1081 LAFTSWMXVTHRCPADVCIELRGAPRKIQTSRVLSLFFNNEPAIGOKLXYTQAAKAANP 1140
QY 1141 GAITVEAGCATTETTTATADARGLIOSSRAHAVALTRHTKCVILDAPGLREVCIT 1200
DB 1141 GAITVEAGCATTETTTATADARGLIOSSRAHAVALTRHTKCVILDAPGLREVCIT 1200
QY 1201 SDVIYNNFFLAGGEVGHRSVIRPCGNPDNLGLTQAFPPSCQISAYHOLAEELGHRPAP 1260
DB 1201 SDVIYNNFFLAGGEVGHRSVIRPCGNPDNLGLTQAFPPSCQISAYHOLAEELGHRPAP 1260
QY 1261 VAAVLPCEPELEBGLLYMPOLELTVSDSVLVELELTDIVHCRMAPRSQKAVLSTLVGRYGR 1320
DB 1261 VAAVLPCEPELEBGLLYMPOLELTVSDSVLVELELTDIVHCRMAPRSQKAVLSTLVGRYGR 1320
QY 1321 RTLYEAASDVRESLARFTPTIGPVQATTCELYELVEANWEKGQDSAVLELDCNRDV 1380
DB 1321 RTLYEAASDVRESLARFTPTIGPVQATTCELYELVEANWEKGQDSAVLELDCNRDV 1380
QY 1381 SRITFFQKXCNKFTTGETTANGKVGOGISAMSKTFCALFEPWRRAIEXETALLPNIIF 1440
DB 1381 SRITFFQKXCNKFTTGETTANGKVGOGISAMSKTFCALFEPWRRAIEXETALLPNIIF 1440
QY 1441 GDATIESVFAAAVSGAGSCWVFENDSEFSTONNNSLGLCECVVMECGMPWLILRLYLH 1500
DB 1441 GDATIESVFAAAVSGAGSCWVFENDSEFSTONNNSLGLCECVVMECGMPWLILRLYLH 1500
QY 1501 VRSAWTLQAPKESLKGFWKHSCEPCTLLMNVTYNNAAITAHCEYEFDFRVAAPKGDSDVY 1560
DB 1501 VRSAWTLQAPKESLKGFWKHSCEPCTLLMNVTYNNAAITAHCEYEFDFRVAAPKGDSDVY 1560
QY 1561 LGSDDYROSRRMAALLIAGCGIKLVDRPGLTYAGVVVAPGLTLPVVPVAPAGLSKKNNG 1620
DB 1561 LGSDDYROSRRMAALLIAGCGIKLVDRPGLTYAGVVVAPGLTLPVVPVAPAGLSKKNNG 1620

[illegible]

Db	70	almleagycrtaraagcclewgahprflndhpnvlnhcfllrprgicdvqwsaprlrpaan	129
Qy	121	CRSALGLPAPRTVYCFDFSRCAFAETGVALYSLHDLPWADVAEAMARHGXTLYAA	180
Db	130	crrsalrglpvdrtycofdgfsrcafaetgvalyslhdlpwadvaeamarhgmrlryaa	189
Qy	181	LALPEVLLPBGTYHTTSTYLLHDDGRANVYTEGOTSAGYNHDSILPAMIRTTKIYGDH	240
Db	190	lhlpewllppgyhtstysyllhdgravvteygdtsagymhdvslrawrltcklygdh	249
Qy	241	PLVIERRALGCHVLLTAPRPSMVPVPRSRSEVVRNITFPGSGPSLFPASASTK	300
Db	250	plvleralagchvllltapepsmpvprsrsevvrsitfpgsgpslfpasastk	309
Qy	301	STFLAVPVHIMDLRMFGALIDDOACCSRCNLTMYLNGISYKVTYGAVALANGMNASBDAL	360
Db	310	stflavpvhimdlrmfгалiddoaccsrcnltmylnglsykvtygalvanegwnasbdal	369
Qy	361	TAXITTAAYLTICHORYLRTQAIKGMRLGVEHNAQFTRLYSWLFPEKSGRDYTPGROLQ	420
Db	370	tavltaaayltichbqyrlrtqaiskgmrlevehaqfllrlyswlfeksgrdytpgrlq	429
Qy	421	FVACCRMLLSGFHLDRRLVVPDESVPBCRRPRLKKVNAKFECCFMRWJGOEOTCLPEAE	480
Db	430	fvaqcrmlsghfldrrlvvpdesvpbcrrprrlkkvnaqfcccfrmwjgoetclpeae	489
Qy	481	GLVGDGHGHDNAYAGSEVSDPAEPANHLDSGTYAVHNOLEALYRLALNPODIARASRLT	540
Db	480	glvgdghgdnayagsevsdpaepeanhlmsgtyavhnolealylrallnppodiaraasrlt	549
Qy	541	ATVELVASPDRLRCRFLVGNKTRFTTYVDGAILLEANGPEQOYVLSFDASROSKGAGSHSLT	600
Db	550	atvelvaspdrlrcrflvgnktrfttyvdgailleangpeqoyvlsfdasrmsgagshslt	609
Qy	601	VELPAGLOVATISSNGIDCATRPPGAPARAAAGEVAACSLALYRNNRTOHSLTGLTW	660
Db	610	yelpaglovatissngidcatrppgaparaaagevaacsalylrynntohsltgltw	669
Qy	661	LHPEGLILTFPPFSPGHIWESANPCGEGTLYLRWTSGFSFSDSPPEAARPAANAAPG	720
Db	670	lhpegllgltfppfspghlweanpfcogegllyrwtsgfsfssdppeaarpnaaapg	729
Qy	721	LPHSTPVPVSDIWLPRPSEEFQVDABLVPRAPDAGLPGCVLVT-PRPPEVPHKPSLPP	779
Db	730	lphstpvpsdiwlvprpseefqvdaablvprapdaglpgcvlvt-prpprevphkpslpp	789
Qy	780	SNRRRLTYTDGAKVYAGSLFESDCDMVYNANPCBRCGGCICAHFYORPEAEATYPRF	839
Db	790	snrrrltytdgakvyagslfesdcdmvynanpcbrcggcicahfyorpeaeatyprf	849
Qy	840	IMRECLAAVYTLPPRIHAHVAPODYRVQONKRLLEAAVRECSRGRGAAAPLGLSGITQVP	899
Db	850	imreclaaytvlpprihaahvapodyrvqonkrlleaavrecsrgrgaaplglsgitqvp	909
Qy	900	VSLSEDAWERNRHRCGDELYLTERAANFEANKRQAPVLTTEDTARTANLALEIDATEV	959
Db	910	vslsedawernrhrcgdeylilteraanfeankrqapvlttedtartanlaleidatev	969
Qy	960	GRACAGCTISGCIYHYQFTAGYPRSGSKSRKIOGDVDVYVPRRELNRMRKRGAAATP	1019
Db	970	gracagctisgciyhyqftagypsgsksrkiogdvdyvprrelnmrkrrgaaatp	1029
Qy	1020	HFAAVTITGRRVVIDEAPSLPRLHLLJHMOASSVHLDPDNPPIADENHGLVPALRP	1079
Db	1030	hfaavtitgrrvvideapslprhlhljhmorassvhlppdnppiadenhglvpalrp	1089
Qy	1080	ELIAPTSMXVYTHRCPADVCELIRGAYKPIQTTSRVLSLFWMERPAIGOKLYVXTOAKAAN	1139
Db	1090	elaptsmwvthrcpadvcelirgaykpiqttsrvlslfwmerpaigoklyvxtoakaan	1149
Qy	1140	PEATVHEAOGATFETTIIATADAKLIDSSRAHAIVALTRTTECVTLDPAGLLREVG	1199
Db	1150	peatvheaogatfettiiataadarglidsrhaahavaltrttecvtlldpaggllrevg	1209

QY	1200	ISDVIYNNFPLAGGEVGHKRPSTVYPRKNPQNIQGTIQAPRPSQISATYIQIALEELGHRPA	1259
Db	1210	ISdviYnnfFlAggevghKhrpsvYprKnPqniQgtIqApRpsQisAtYiqIAleELghRpa	1269
QY	1260	PVAANLPPCPPELEOGLLYMPDELTSVSVLYFELTQIHVCRMNAPSGRKAVLSTLYGRY	1319
Db	1270	pvaANlppcPeLeogllYmpDeltsvsvlyfELtqiHvCRmNAPsGRkAVlSTLYgRy	1329
QY	1320	RRTKLYEAHSDVBESLARFIPTIGPVQATTCELYELVEMVKEGQDGSAYLELDLCNRD	1379
Db	1330	rRtklyeaHsdvBeslarfIptIgpvQattCeLyElvEmvKEgQdGSaYleldLCnrD	1389
QY	1380	VSRTTFPPKXKXNKTGTGTTAHGKVGCGISANMKTSCALFGFMFRAIEKEILLPPNIF	1439
Db	1390	vSrttFpPKxKXnKtGTgTtAhGkVgCGisAnMkTscALFGFMfRAIEkEILLppNif	1449
QY	1440	YGDAAVEESVFMAAASGAGSCGVFENDFSEFDSRONNFSLGLEPCVYMEECGMPQWILRLYH	1499
Db	1450	yGdaaVEesVFMAaASgAGsCvFenDFsefDSrONnFslGLEpCvYMEeCGmpQwILrlyH	1509
QY	1500	LYRSAMWIIQAPKESLKGPMKKHSGEGECLLMNTVMMAITAHCYERDFRVAARFGDDSV	1559
Db	1510	lyrsAmwIIQApKESlKgPMkKHSGEGeCLlMntVMmAItAHcYERdFRvAARfGDDsv	1569
QY	1560	VLCSPYRORSNAAALLINGCGIKLKVDYRPGLAGVYVAPRGITLPDYVRFAGRLSEKNW	1619
Db	1570	vLcspYrORSnAAaLLIngCGIKlkVdYrPglAGvYvAPrGiTLpDYvRFaGRlSEKnw	1629
QY	1620	GGPERAEQRLAACDRLRGITVNAVQCVDVVSRVGVSPGLVHNLIGMLQTIADKAHF	1679
Db	1630	gGpErAEqrlAAcDRLrgITvNAVqCvDVvSRvGVSPgLVhNLIGmLQtiADKAhf	1689
QY	1680	TEITKPVLDLTNSITQKVE 1698	
Db	1690	teItKpVldLTnSiTqKve 1708	
RESULT	3		
VJ1384			
ID	VJ1384	standard: Protein: 1708 AA.	
XX	XX		
XX	YJ1384;		
AC	XX		
DT	12-Oct-1999	(first entry)	
XX	XX		
DE	HEV-US2 ORF1 protein.		
XX	XX		
KW	Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;		
KW	vaccine; passive immunisation.		
XX	XX		
OS	Hepatitis E virus.		
XX	XX		
FH	Key	Location/Qualifiers	
FT	Misc-difference 322	/label= unknown	
FT	FT	/note= "encoded by CRG"	
FT	Misc-difference 331	/label= unknown	
FT	FT	/note= "encoded by RAC"	
FT	Misc-difference 445	/label= unknown	
FT	FT	/note= "encoded by RAC"	
FT	Misc-difference 448	/label= unknown	
FT	FT	/note= "encoded by RAC"	
FT	Misc-difference 634	/label= unknown	
FT	FT	/note= "encoded by RTG"	
FT	Misc-difference 646	/label= unknown	
FT	FT	/note= "encoded by CCY"	
FT	Misc-difference 811	/label= "encoded by GCS"	

FT	/label= unknown	
FT	/note= "encoded by TKT"	
FT	1533	
FT	Misc-difference	
FT	/label= unknown	
FT	/note= "encoded by TAY"	
FT	1578	
FT	Misc-difference	
FT	/label= unknown	
FT	/note= "encoded by RGC"	
FT	1691	
FT	Misc-difference	
FT	/label= unknown	
FT	/note= "encoded by GAR"	
PN	WO9919732-A1.	
PD	22-APR-1999.	
XX		
PF	15-OCT-1998;	98MO-US21941.
XX		
PR	15-OCT-1997;	97US-0061199.
XX		
PA	(ABBO) ABBOT TM LAB.	
XX		
PI	Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;	
XX		
DR	WPI; 1999-288017/24.	
DR	N-PSDB; 200267.	
XX		
PT	Detection of United States isolates of hepatitis E virus	
XX		
PS	Claim 6; Page 216-221; 260pp; English.	
XX		
CC	The invention provides a method for detecting a US (sub)type hepatitis E	
CC	virus (US-HEV), or its naturally occurring variants in a sample by	
CC	treatment with a binding partner specific for a marker of the virus, and	
CC	then detecting any complex formed. The method is used to diagnose	
CC	infection with US-HEV. Polypeptides from US-HEV, antibodies specific for	
CC	open reading frames (ORF) in US-HEV and host cells expressing these ORFs	
CC	are useful in vaccines or for passive immunisation. The polypeptides are	
CC	also used to raise specific antibodies (useful as immunoassay reagents).	
CC	Fragments of nucleic acid from US-HEV are useful as primers and probes in	
CC	usual hybridisation and amplification assays for detecting infection. The	
CC	present sequence represents a HEV-US2 ORF1 protein.	
XX		
SQ	Sequence 1708 AA;	
	Query Match 97.1%; Score 8753.5; DB 20; Length 1708;	
	Best Local Similarity 97.2%; Pred. No. 0;	
	Matches 1651; Conservative 8; Mismatches 39; Indels 1; Gaps 1;	
OY	1 PEITTAIEQALAAANSALANAVVVRPELSRVOTEILINLMOPROLVREPEVLNMHPIOR	60
DB	10 pgitatlaqaaalaanaalanavvrpfisrvqteailinlmprqlvfrpevlmhpiqr	69
OY	61 VHNNELEQYCRARAGRCLEVGAHPRISINDNPVLRHRCFLRPVGRDVRORYSAFPRGPAN	120
DB	70 vhnneleqycraragrclevgahprisindnpvlrhcflrpvgrdvrwysaprtgpaan	129
OY	121 CRRSALRGLPADRTYCFDGSFRCAPFAAETGVALYSLHDLMADVAAEAMARRGXTRLYAA	180
DB	130 crrsalrglpadrtycfdgfsrcafaetgvalyslhdlmadvaaeamarrgmrtlyaa	189
OY	181 LHLPEVLLPGTYTTSYVLLIHODRAVVMYEGOTSAAGYNNDVSTLAAWIRTKTYGDH	240
DB	190 lhlpevllpgtyttsyvlilhodravvmylegotsagynndvslawirtktygdh	249
OY	241 PLVLEKVAIAICHEFULLTAAPESPMYVVRPRSTEVYVVISIFEPGSGSPSLFPACSTK	300
DB	250 plvlekvaiacchefulltaaapespmvyvvrprstevyvisifpgsgspslfpacstk	309
OY	301 STFAAVPVHINDRLMFCATLDDQAFCCSRMLTYLGRISYVYVYGALVANEGWNASDAL	360
DB	310 stflavpvhndrlmfcatlxdqafccsrmltylgrisyvvtvgaivanegwnasdal	369

QY 361 TAAITAAVLTICHQRYLRQAIISKGRRLGVEHAOKETITRLYSWLFPEKSGRDYIPGRLO 420
 Db 370 TAVITAAVLTICHQRYLRQAIISKGRRLGVEHAOKETITRLYSWLFPEKSGRDYIPGRLO 429
 QY 421 FYAACRRLMSAGFHLDPRLVLPDESVPICRQKRLKRVAGKCCFMRMLGECFCFLEPAE 480
 Db 430 FYAGCRRVLSAGFHLKPRXIVFDESVPICRQKRLKRVAGKCCFMRMLGECFCFLEPAE 489
 QY 481 GLVGDHGHNEAVEGSEVDPAPRAHLDVSGTYAVHGHOLEALYRALNPODITAKRSRLT 540
 Db 490 GLVGDHGHNEAVEGSEVDPAPRAHLDVSGTYAVHGHOLEALYRALNPODITAKRSRLT 549
 QY 541 ATVELASPDRLCETRYVLGKTRFTYVDOGAHLIANGPEQYVLSFPAASRQSMAGSHLT 600
 Db 550 ATVELASPDRLCETRYVLGKTRFTYVDOGAHLIANGPEQYVLSFPAASRQSMAGSHLT 609
 QY 601 YELTPAGLVRISSNGLDCTATPPPGAPSAAGAEVAFCALRYNRFTQRHSLNGLM 660
 Db 610 YELTPAGLVRISSNGLDCTATPPPGAPSAAGAEVAFCALRYNRFTQRHSLNGLM 669
 QY 661 LHPGGLGTPPFPSPGHIMESANPECGEGLYRTWSTSGSSDFSPPEAAPAMATPG 720
 Db 670 LHPGGLGTPPFPSPGHIMESANPECGEGLYRTWSTSGSSDFSPPEAAPAMATPG 729
 QY 721 LPHSTPPVSDIWLPPRSEFQVDAAPVPPAPAGLPGVLT-PPPPPPVKKPSIPP 779
 Db 730 LPHSTPPVSDIWLPPRSEFQVDAAPVPPAPAGLPGVLT-PPPPPPVKKPSIPP 789
 QY 780 SRNRRLLYTPDGAKYAGSLFESDCDMLVNASNPGHRCGGCHAFYOEFPAPFPTF 839
 Db 790 SRNRRLLYTPDGAKYAGSLFESDCDMLVNASNPGHRCGGCHAFYOEFPAPFPTF 849
 QY 840 IMREGIAVTLTPRPILHAAVADRYEQNKRLAAVREFCSRRAAYPLLSGSIQYV 899
 Db 850 IMREGIAVTLTPRPILHAAVADRYEQNKRLAAVREFCSRRAAYPLLSGSIQYV 909
 QY 900 VSLSPFAMENHRRPDELITTEPAANMFENAKRAQVLTTEETARANALALIDAATEV 959
 Db 910 VSLSPFAMENHRRPDELITTEPAANMFENAKRAQVLTTEETARANALALIDAATEV 969
 QY 960 GRACACTISPGIVHYOFTRGVPGSGKRSIQGDVDVVPVRELNSRRRGFAATP 1019
 Db 970 GRACACTISPGIVHYOFTRGVPGSGKRSIQGDVDVVPVRELNSRRRGFAATP 1029
 QY 1020 HFAARVTIGRRVVIDEAPSLPRLHLLHMQRASSVHLGDPNOIPALDFHAGLVPAIRP 1079
 Db 1030 HFAARVTIGRRVVIDEAPSLPRLHLLHMQRASSVHLGDPNOIPALDFHAGLVPAIRP 1089
 QY 1080 ELAPTSMAWYTHRCPADVCELRGATPKIOTSRVLSLFWMNPAIGOKLVXIOAKAAN 1139
 Db 1090 ELAPTSMAWYTHRCPADVCELRGATPKIOTSRVLSLFWMNPAIGOKLVXIOAKAAN 1149
 QY 1140 PGATVTHEAGATFETTTIATADAGLTIOSRAHAIVALTRHTEKVTIADAGLREVG 1199
 Db 1150 PGATVTHEAGATFETTTIATADAGLTIOSRAHAIVALTRHTEKVTIADAGLREVG 1209
 QY 1200 ISDVIYNNEFLAGGEVGHXRPVIRPGNDPONLGTLOAPPSQOISAYHOLAEELHHPA 1259
 Db 1210 ISDVIYNNEFLAGGEVGHXRPVIRPGNDPONLGTLOAPPSQOISAYHOLAEELHHPA 1269
 QY 1260 PVAANVLPCELEGOGLIYMQELTVSDVLYFELTDIVHCRMAAPSGRAKAVLSTLVGRY 1319
 Db 1270 PVAANVLPCELEGOGLIYMQELTVSDVLYFELTDIVHCRMAAPSGRAKAVLSTLVGRY 1329
 QY 1320 RRTKLEAASDVRESLARIPTIGVQATTCLELYEAMVEKGGDGSVLELDCNRD 1379
 Db 1330 RRTKLEAASDVRESLARIPTIGVQATTCLELYEAMVEKGGDGSVLELDCNRD 1389
 QY 1380 VSRITFPOKKXCNKFTTGERTIAHGRVGGISAMSKFTCALFGPWFRAIEKETALLPNNIF 1439
 Db 1390 VSRITFPOKKXCNKFTTGERTIAHGRVGGISAMSKFTCALFGPWFRAIEKETALLPNNIF 1449

QY 1440 YGDATSEVFAAASGAGSCAMFENDESPDSTQNNFSLGLECVMECGMPQMLRLTYH 1499
 Db 1450 YGDATSEVFAAASGAGSCAMFENDESPDSTQNNFSLGLECVMECGMPQMLRLTYH 1509
 QY 1500 LVBSAMILQAPKESLKGFMKHSGEPTLLMNTVMNAIILAHCEYERDFRVAAFKGDSDV 1559
 Db 1510 LVBSAMILQAPKESLKGFMKHSGEPTLLMNTVMNAIILAHCEYERDFRVAAFKGDSDV 1569
 QY 1560 VLCSVDYRQSNAAALLAGCGLAKVDYRPIGLAGVVAAGLTDLDVYVAFGRLSKKN 1619
 Db 1570 VLCSVDYRQSNAAALLAGCGLAKVDYRPIGLAGVVAAGLTDLDVYVAFGRLSKKN 1629
 QY 1620 GPGPERAEOLRLAVCDPLRGLTNVAQCVDSRVGSPGLVHNLIGMLQTTADGKAHF 1679
 Db 1630 GPGPERAEOLRLAVCDPLRGLTNVAQCVDSRVGSPGLVHNLIGMLQTTADGKAHF 1689
 QY 1680 TETIKPVLDTNLSITQKVE 1698
 Db 1690 TETIKPVLDTNLSITQKVE 1708
 RESULT 4
 R91813
 ID R91813 standard; Protein: 1693 AA.
 AC R91813;
 DT 26-NOV-1996 (first entry)
 XX
 DE Hepatitis E virus strain SAR-55 ORF-1.
 XX
 KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
 KW structural region; antigen; detection; antibody; vaccine;
 KW immunisation; infection.
 XX
 OS Hepatitis E virus.
 XX
 FH key Location/Qualifiers
 FT Misc-difference 1238
 FT /note- "corresponding codon CAG"
 FT Misc-difference 1455..1693
 FT /note- "10 bp nucleic acid sequence TGGTNTTTCGA
 FT has to be inserted between nucleotides
 FT 4390..4391 of T27394 before these amino
 FT acid residues can be decoded"
 FT
 W09610580-A2.
 PD 11-APR-1996.
 XX
 PF 03-OCT-1995; 95WO-US13102.
 XX
 PR 03-OCT-1994; 94US-0316765.
 XX
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 PI Emerson SU, Purcell RH, Tsarev SA.
 XX
 DR WPI: 1996-209320/21.
 XX
 DR N-PSDB: T27394.
 XX
 PT Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
 PT antigenic protein useful in diagnosis, prophylaxis and treatment of
 PT hepatitis E virus infection
 XX
 PS Disclosure: Pages 9-13; 121pp; English.
 XX
 CC The present sequence is the protein prod. of ORF-1 from the
 CC hepatitis E virus (HEV) strain SAR-55, which was implicated in an
 CC enterically transmitted non-A, non-B hepatitis in Pakistan. The
 CC protein encoded by the structural region of the virus (i.e. ORF-2),
 CC which is capable of forming HEV like particles, is useful for the
 CC detection of HEV antibodies (pref. IgG or IgM) in blood, plasma,

CC sera, cerebrospinal fluid, tissue, urine or pleural fluid. The
 CC protein, and anti-HEV antibodies generated using the protein, can
 CC also be used in vaccines for immunising an animal against HEV
 CC infection. The protein is identified as a band of greater than
 CC 50 kD following SDS-PAGE of cell lysates of insect cells infected
 CC with a HEV ORF-2 contg. baculovirus, i.e. the claimed recombinant
 CC expression vectors pPIC9-1779, -1780 and -1781.

XX Sequence 1693 AA;

Query Match 82.7%; Score 7456; DB 17; Length 1693;
 Best Local Similarity 81.8%; Pred. No. 0;
 Matches 1400; Conservative 110; Mismatches 160; Indels 42; Gaps 7;

QY 1 PGITTAIQALAAANSLANAVVPRFLSRVQTEILINLMQPROLYRPEVLEMHPIQR 60
 Db 10 pgitatqaaalaanaaanaavvrflsqhqllellnlnmqprlvtrpevfwhpqr 69
 QY 61 VIHNELEOYCARACRCLEVGANHPISINDPNVLRHCLRPVGRDVRKWSAPTRGPAN 120
 Db 70 vihnelelycraargrclelghprslndpnvhrclrpagrdvqrwytacltrgpaan 129
 QY 121 CRRSALRGLPADRTYCDGFSRCAFAETGVALYSLHDLMPADVAEAMARHGXRLYAA 180
 Db 130 crsallrlypadrtlycdgfsrcafaetgvalyslhdmgsdvaeamfrhgmrllyaa 189
 QY 181 LHLPEVLLPFGYHTTSTYLLIHDDRAVVVTEGDSAGYNDVSIILRWIRTKTKVD 240
 Db 190 lhlpevllppgytlaayllihdgrlvvteyegdsagynhdvsnlswirtekvdh 249
 QY 241 PLVIERVAIGCHFYLLTAAPESPMPYVPRSTEVYNSIFPGSGSPSLFPACSTK 300
 Db 250 plviervaligchfylltaapepmpypvprstevynsilfpggslpslfpstcsk 309
 QY 301 STFNAPVPHIMDLMLFGATLDDQAFCCSRMLTYSYKYTVGALVANEGRMNSDAL 360
 Db 310 stfnavphimdlmlfgatlddqafccsrmltlyrlyskvtyglvaneegmasedal 369
 QY 361 TATITTAALYTCHORLYRTOALSKGMRLGVEHAKFTIRLYSWLEFSGRDIYIPGRLO 420
 Db 370 tavitlaalytlchqyrltgalskgrmlerehaqkfltrlyswlfeksgrdyipgrql 429
 QY 421 FYACCRRLSAGFHLDPRVLYFDESVPQCRFTFLKAVGRCCEPMRWLGQECTCLEAE 480
 Db 430 fyaccrrltsagfhlndprlvlyfdesapchrtalrkvsktctfkmkygqectcllqgae 489
 QY 481 GLVGDHGDHNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALYPODIAARSLT 540
 Db 490 glvgddqghdneaeyegsdvpaesaidslsgsyvvpqtalqplqgaldlpaetlvaagrlt 549
 QY 541 ATVELVASPDRLECRVLYGNKTFRTTVVDGAHLEANGPEOYVLSFASRQSGMGSLSLT 600
 Db 550 atvsvsqvdrlecdvlygnkftfsvdgvleungperhmsfdaesqsmagrsfslt 609
 QY 601 YELTPAGLOVRISSNGLDCTATFPPGAPSAAPGEVAFCALRYKNFTORHSLTGLWM 660
 Db 610 yaasaaglevryvaagldhrtavfagvpsrzapgevtalcyfynreaqrslstlgnfw 669
 QY 661 LHPGGLGIFPPFSGHIMESANPFCGGLTYTRTW-----TSGESSDFSPPE 709
 Db 670 fhpegllipflapfispghwesapfcgsclylrltwesvdavspapqdlgfts-----e 724
 QY 710 AAAPAMATPGLPHSTPVSIDWLPPSEEFQYDAAPV-PPADPA-GLPGPVLYLPP 767
 Db 725 psipratcp-----tp-----aaipppadpspsltsapargepap 761
 QY 768 PPVYHAKSIP-PPSRNRRLTYTPDGAKVYAGSLFESQCDMLVNASNGHRRPGGLCHAF 826
 Db 762 gataarapaltqatthrllfitypdskvfagslfestctvlnasnvdhppgslchaf 821
 QY 827 YORREAFYPTFEFIMRGSLAYTLTPRIITAAVAPDVRVEONPKRLAAYETCSRKETA 886
 Db 827 yorrefafyptfefimrgslaytltpriitaaavpdrvveonpkrlaayetcsrketa 886

Db 822 yqrypasfdaesfymrdgaaytltpripilhavapdyrlenhpkrlaayretcsrlgta 881
 QY 887 APLILGSGIYOVPVLSFSDAMERNHRPDELYLTPPAANWFEANKPAQVLTITPDRRT 946
 Db 882 aypilgsgiyovpvlstsdameernhrpdeyllytpaarnwfeanpqaqvltitpdrart 941
 QY 947 ANLALEIDATEVGRACAGCTISPGIVHYOFTAGVPSGSKSRISIOGDDVYVVPTRRLR 1006
 Db 942 anlaleidsatdvgracagcvtlpgvvyqftagvpsgsksrslsqadavdvvprrlrl 1001
 QY 1007 NSMRRGFAATPHTHAAVTITGRRVVIDEASLPRLHLLLMQRRASSVHLLGDPQIPAI 1066
 Db 1002 nsmrrgfaatphtthaavtigrvvideaslprrllhllmqrraatvhlldgdpqipai 1061
 QY 1067 DEEHAGVPAIRPELAPRSMXVTHRCPADVCELRGAYPKIOTTSRYLRSFENNEPAIG 1126
 Db 1062 deehagvpaairpelaprsmxvthrcpadvcelrgaypkioettsrylrsfennepaig 1121
 QY 1127 QKLYVTOAKAANPGALITVHEAGATFETTTIATADARGLIOSSRAHAYALTRHTEKC 1186
 Db 1122 qklyftgaakaanpgsvltvheagatyfettliatadargliqssrahaivaltrhtkc 1181
 QY 1187 VILAPGLLRREVGISDVIVNNPFLAGGEVGHRSVTPRGMPDONLGTLOAPPSCOISA 1246
 Db 1182 vilapglrrrevgisdvivnnpflaggevgghrsvtprgmpdonlgtloapppscois 1241
 QY 1247 YHOLAEELGHRPAPVAAVLPCCPELEOGLLYMPOELTVSDSVLYFELTDIYHCRMAAPSQ 1306
 Db 1242 fhqlaeelghrppapvaavlpccpeleqlllylpqelctcdsvvtfeldivhcrmaapsg 1301
 QY 1307 RKAVLSTLVGRYGRRTKLYEAHSDVRESLARFPTTPVQATTCELLYELVANEVKGOD 1366
 Db 1302 rkavlstlvgryrtrtklyeahsdvreslarfpttpvqatccellyelvanevkgod 1361
 QY 1367 GSAVLELDLCNRDVSRIEFPKXCNKFTTGTEIHAHGKGGCISAMSKFCALFGWPFPAI 1426
 Db 1362 gsavleldlcsrdvsrilefkdckkfttgteiahgkvvggisaaskfcaltgwpfpa 1421
 QY 1427 EKEITALLPNIIFYGDAYEESVFAAASVAGSCEWFEENDFSEPDSTONNFSLGECVME 1486
 Db 1422 ekaitallpnyifygdadtdvfaaavaakasmfendfsefstqnfalsgecalme 1481
 QY 1487 ECGMPQHLIRLHLVRSWMLIQARKESLKGWKKHSGSPGLNNTVNMALIAHCYEF 1546
 Db 1482 ecgmpqhllrhlhlrseswmlqarkeslrgfwkkhsgspglntvnmavilchcyef 1541
 QY 1547 DFRVAEFKDDSVVLCSDYROSRNAALIIACGGLKLKVDYRPIGILYAGVVAAPGIGTLPD 1606
 Db 1542 dlqvaaefkgddsvlylcsdyrsgpaavliagcgjklkvdfrpildlyagvvaapigalt 1601
 QY 1607 VVRPAGRLSEKNMGPGPEPABOLRLAVCDFLRGLTNVAQVCVDVVSRYGVSPGLVHNL 1666
 Db 1602 vvrpagrllseknmgpgpepaeqlrlavsdflrkltnvaqvmcvdvvsrygvspglvhn 1661
 QY 1667 GMLDTIADGKAHFTETIKPVLDLTNSIIORVE 1698
 Db 1662 gmladvdgkahtesvykpvldltlnsllicrve 1693

RESULT 5
 W80196
 ID W80196 standard; Protein; 1693 AA.
 XX
 AC W80196;
 XX
 DT 23-DEC-1998 (first entry)
 XX
 DE Protein encoded by ORF1 of ET-NANB (HEV) Burma strain DNA sequence.
 XX
 KW Enterically transmitted non A non B hepatitis virus; ET-NANB;
 KW Hepatitis E virus; HEV; Burma HEV isolate; vaccine;
 XX diagnostic probe.
 XX

OS Non A non B Hepatitis virus.
 XX US5824649-A.
 PN 20-OCT-1998.
 XX 07-JUN-1995; 9505-0475807.
 PF 25-JUL-1994; 9405-0279823.
 XX 17-JUN-1988; 8805-0208997.
 PR 11-APR-1989; 8905-0336672.
 PR 16-JUN-1989; 8905-0367486.
 PR 13-OCT-1989; 8905-0420921.
 PR 05-JUL-1990; 9005-0505888.
 PR 07-JUN-1995; 9505-0475807.
 XX
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PI Bradley DM, Fry KE, Krawczynski KZ, Reyes GR, Tam A;
 PI Yarbough PO.
 XX WPI: 1998-582599/49.
 DR N-PSDB: V66321.
 XX
 PT Hepatitis E virus proteins - useful for diagnosis or vaccine
 PT production the virus
 PS
 XX Claim 22: Columns 57-66; 47pp; English.
 CC W80196-98 are encoded by the genome of the Burma strain of
 CC enterically transmitted non A non B hepatitis virus (ET-NANB)
 CC (hepatitis E virus (HEV)). The specification describes an isolated
 CC protein which is specifically immunoreactive with antibodies present
 CC in individuals infected with HEV and encoded by a sequence contained
 CC in an open reading frame (ORF) of an HEV genome. The genome has a
 CC sequence that is more than 70% identical to the ORF1 sequence from
 CC Burma HEV isolate. The protein is used as a vaccine and a diagnostic
 CC probe for ET-NANB.
 XX
 XX Sequence 1693 AA:
 SQ
 Query Match 82.6%; Score 7448; DB 19; Length 1693;
 Best Local Similarity 81.8%; Pred. No. 0;
 Matches 1396; Conservative 110; Mismatches 169; Indels 32; Gaps 6;

QY 421 FVACRRRLISAGPHLDPRVLYPEDESVPCCRRTFLKRVAGKFCFCFMRWLGOECTCFEPAP 480
 |||||
 DB 430 fyaqcrwlsagfhldprylvlpeesapchrtalrkalskfcicmkiqecclfpaae 489
 |||||
 QY 481 GLVGDHGHNEAESEGVPAEPAHLDVSGTVVHGHOLEALRYALNVODIARAARSRLT 540
 |||||
 DB 490 gavgdghndeaegsdpbaesaidsisgyvvpqalpyqaldlpaelvaargrlt 549
 |||||
 QY 541 ATVELVASPDRLRCRTVLGNKTRFTTVVDGAHLNANGPEQVYLSFPDASRQSMGASHSLT 600
 |||||
 DB 550 atvksqvgdrlcdcelnktfritsfvlgavletngpethnlsfdaqslnaagfsilt 609
 |||||
 QY 601 YELTPAGLOYRISNGLDCTATPPPGGAPSAAAEVAALRYNRRTQRSLTGILM 660
 |||||
 DB 610 yaasaaglevryaaagldhnavfapvpsrpsagvafalsaylrlnreaqrhslnglwt 669
 |||||
 QY 661 LHPEGILGIFPPSPGHWESANPFCGEGTLRYRTWS-TSGFSSDSP-----PEAAPA 714
 |||||
 DB 670 fhpegllgltapfspgnwesapfcgescllyrtvsevdavsparpdlgfmspsyps 729
 |||||
 QY 715 MAATPGLPHSTPPSDIWLPPSEEFQVDAFV-PPAPDPAGLP-GPVVLTAPPPEPVH 772
 |||||
 DB 730 raatppl-----aapllpppappspppapalaepasgatag 766
 |||||
 QY 773 KPSIP-PPSRNRRLTYTPDGAAVYAGSLFESPCDMLVNASNGHRPGGGLCAFYQRP 831
 |||||
 DB 767 apaltqhtatrhrltlftypgskvfaslfesctclvlnasnvdhprggglchaifgryp 826
 |||||
 QY 832 EAFVPEFIMREGLAAYTLPRPIHAHAVAPDYVEONPKRELAAYETCSRRGTAAVPL 891
 |||||
 DB 827 asfdaasfmrtdgaayclprplihavadytlempkrlleaayetcsrlgtaaypl 886
 |||||
 QY 892 GSGIYQVPSLSDAEMERNRPDELYLEPAANWEANRPAQVLTITEDTARTANAL 951
 |||||
 DB 887 gtlgyqplpgrdsdawnhrpddelylpelaarfeanrptlilttedvartanal 946
 |||||
 QY 952 EIDATEVGRACAGCTISPEIVHYQFAGYPSGSKSISIOGSDVDVYVTRRLRMSMR 1011
 |||||
 DB 947 eldsatdvgracagcrrvtpgvqyftaagvysgsksrslqadavdvvpvltelrnavir 1006
 |||||
 QY 1012 RGFAAFTPHTAARVTIGRRVVIDEAPSLPHLLLHMQRASVYHLGDDPQIPADIFEHA 1071
 |||||
 DB 1007 rgfaaftphtaarvtgrrrvvdeapslphlllhmgraatvhljgdpnqpaideha 1066
 |||||
 QY 1072 GLVPALRPDLAPTSMAWVTHRCPADVCELIRGAYPKIOTTSRYLSLFENPEAIGOLVX 1131
 |||||
 DB 1067 glvpalrpdlgpsawvthrwpadvcellrgaypmqttsrylsrftwgepavgqlvf 1126
 |||||
 QY 1132 TOAKAANPCATIVVHEAGATFETTTIATADABGLIOSSRAHAYALVTHTEKCYILDA 1191
 |||||
 DB 1127 tgaakpampgsvsvcheagatyletllatacadargllqssrahaivaltlhtekcvilda 1186
 |||||
 QY 1192 PGLIREVIGISDVIVNNEFLGGEVGHXRPVIRGNPDOLGTLQAPPSCOISAVYHQA 1251
 |||||
 DB 1187 pglirevgisdvlnnmflfaggeighqrpvripmpndanvdclaaipscqslafhqla 1246
 |||||
 QY 1252 EELGHRPAAVAVLPCCPELEOGLLWMPQELATYSDSVLVELNDIVHCRMAAPSORKAVL 1311
 |||||
 DB 1247 eelghrnpvaavlpccpellegllylpqelctcdsvtelclidvncmaapsqravl 1306
 |||||
 QY 1312 STLVGKRGRTKYLEAASHSVRESLARFPTIGPVQATTELYELVAMWEKQDGSAYL 1371
 |||||
 DB 1307 stlvgrgygrtklynashsvrdsrlarflpavpyvtlcllyelvaamwexqddgsav 1366
 |||||
 QY 1372 ELDDCNRDVSRITFEQXCKFTTGERTIANGKVGOGISANSKTFCAFGMPRAIEKEIL 1431
 |||||
 DB 1367 elddcnrdvsriltfegdkckfttgetiangkvqgslsaasktfcalfgpwaiaekail 1426
 |||||
 QY 1432 ALLPPIFYGDAYEESVFAAASVAGAGSCWVFENDESEFDSSTONNFSLGLECVYMEEGMP 1491
 |||||
 DB 1427 allppgyfygdarddvtsaavaaaksmvfendfsefdsctgnfnslgcalameegmp 1486
 |||||

QY 1492 CMLILYHLVRSAMLLQAPKESLKGFWKHSCEPGLLMNTVMNMAIIACHYEFDERVA 1551
 DB 1487 qwlrllyhllylrsawllqapkeslrgfwkhsgeplllmwtvmnmaivlthcydfidfyva 1546
 QY 1552 AFKGGDSVLCSDYQSRNMAALGACGLKLVDRPRLIGLVGVVAVPGLGTLPPVVRPA 1611
 DB 1547 afkgddsvlcsdyqsrnmaalgiagcgklkvdrpirlglvgyvvavpglgaltppvvifa 1606
 QY 1612 GRUSKNNMGCPGERAEQLRLAVCDFLRGLTNNVAQVCVDVSVFVSPGLVHNLGMLQT 1671
 DB 1607 grltekngvpgpgeraeqlrlavscdfllrglttnvaqvcvdvsvfsvpglvhnlglmlqda 1666
 QY 1672 IADGKAHFTERTIKPVLDTLNSTIORVE 1698
 DB 1667 vadgkahftesvkpyldltnstiorve 1693
 RESULT 6
 B24119
 ID B24119 standard; Protein; 1693 AA.
 XX B24119:
 AC B24119:
 DE 29-JAN-2001 (first entry)
 XX
 DE Hepatitis E virus Burma strain ORF1 protein sequence SEQ ID NO:7.
 XX
 KW Hepatitis E virus; HEV; non-A/non-B viral hepatitis; viral protein;
 KM immunological; diagnosis; hepatitis; infection; identification;
 KM detection; immunoreactive; hepatotropic; antiinflammatory; virucide;
 KW vaccine; antiviral; antigenic; antibody; antigen.
 XX
 OS Hepatitis E virus.
 XX
 PN US6120988-A.
 XX
 PD 19-SEP-2000.
 XX
 PF 07-JUN-1995; 950S-0478507.
 XX
 PR 25-JUL-1994; 940S-0279823.
 PR 05-APR-1991; 910S-0681097.
 PR 17-JUN-1988; 880S-0208997.
 PR 11-APR-1989; 890S-0336672.
 PR 16-JUN-1989; 890S-0367466.
 PR 13-OCT-1989; 890S-0420921.
 PR 05-APR-1990; 900S-0505888.
 XX
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 PA (USCO) US GOVERNMENT.
 XX
 PI Yarbough PO, Krawczynski KZ, Fry KE, Bradley DW, Tam A, Reyes GR;
 XX
 DR N-PSDB; A55259.
 XX
 PT MPI: 2000-593712/56.
 PT
 PT Identifying recombinant antigen immunoreactive with antibody induced by
 PT hepatitis E virus (HEV), for detecting HEV infection, comprises
 PT immunoreacting a polypeptide from an HEV genome with an HEV-positive
 PT antiserum -
 XX
 PS Claim 6; Column 55-64; 46pp; English.
 XX
 CC The present invention describes a method for identifying a recombinant
 CC antigen immunoreactive with a hepatitis E virus (HEV)-induced antibody.
 CC The method comprises producing a polypeptide derived from an HEV genome,
 CC immunoreacting the polypeptide with an HEV-positive antiserum and
 CC selecting the polypeptide as a recombinant antigen if the polypeptide
 CC reacts with the HEV-positive antiserum. The method is useful for
 CC identifying recombinant antigen immunoreactive with antibody induced by
 CC HEV. The enterically transmitted non-A/non-B hepatitis (ET-NANB (also
 CC known as HEV))-specific fragments are useful for identifying ET-NANB-
 CC derived cDNAs, which contain additional sequence information, as primers

CC for detecting ET-NANB viral genomic material in a patient sample, for
 CC the synthesis of polypeptides for use in immunoassays, and for
 CC identifying similar antigenic regions encoded by related viral strains,
 CC e.g. Burmese strain. The antigens are especially useful in the
 CC preparation of vaccine against ET-NANB infection. These antigens may
 CC further be used to prepare antibodies to ET-NANB virus particles for
 CC pre- or post-exposure prophylaxis. The present sequence represents a
 CC specifically claimed HEV Burma strain protein sequence for use in the
 CC present invention.
 CC
 XX
 SO Sequence 1693 AA:
 Query Match 82.6%; Score 7448; DB 21; Length 1693;
 Best Local Similarity 81.8%; Pred. No. 0;
 Matches 1396; Conservative 110; Mismatches 169; Indels 32; Gaps 6;
 QY 1 PGITTAIOAALAAANSALANAVVVRPPLSRVOTIELINLMQPROLVFREPVLNHP1QR 60
 DB 10 pgitatagaalaansalanavvrplsrvtielinlmqprqlvfrpevfwnhp1qfr 69
 QY 61 VIHNELEQYCRARAGRCLEVGARHPRSTINDNPNVLHRCFLRPVGRDYOQWYSAPTRGPAAN 120
 DB 70 vihnelleqycrarsgrclevgahprstindnnpvnlhrcflrpvgrdvyqwtysaptrgpaan 129
 QY 121 CRRSALRGLPPADRTYCFDFGFSRCAFAETGVALYSLDHPADVAEAMARHGXTRLYAA 180
 DB 130 crrsalrglppadrtycldfgfsgcnfaetgialysldhmpsdvaeamfthgmtrlyaa 189
 QY 181 LHLPEVLLPRTYHTTYSYLLIHGDRRAVVTYEGDTSAGVNHDSYILRAWIRTKIYGDH 240
 DB 190 lhlpevllprtyhttyyllihgdravvtyegdtsagvnhdsvilrawirtkivgdh 249
 QY 241 PLVIERVAICGFVLLTLTAPESPMPYVPRSTEVYVNSITRGSGSEPLFSASCTK 300
 DB 250 plviervaicghvllltaapepsmpyvyprstevyvsitrgpgtspilfptsctk 309
 QY 301 STFAVAVPHIMDRMLFGATLDDOAFCCSRMLTYLRGISTYKVYTGALVANAGMNASEDAL 360
 DB 310 stfnavpahimdrmlfgatlddgaefccsrmltylrgistyvtygtlvlanegmnaasedal 369
 QY 361 TAXITTAAYLTCHORYLTQALISKGMRLGVEHAKFTRLYSWLPEKSGHDYIPGRLOQ 420
 DB 370 taxitaayltchqrylrltqalskgmrllerehagkftrlyswlfeksgdyipgrgle 429
 QY 421 FYACCRHWSAGFHLDPRLVLFEDSVPCRCRFTFKKVAAGKCCMRMLGOCCELTAPAE 480
 DB 430 fyaccrhwslagfhlprlvlfedsvpcrcrftfkvvaagkccmrmlgoccteltapae 489
 QY 481 GLVGDHNDNEAYEGSEVPAEPAPHLDVSGTYAVHGOLEALYRALVNPODIARASRLT 540
 DB 490 glvgdghndneayegsdvdpesaisdlsysgyvpygalqlpialdlpeilvaragrlt 549
 QY 541 ATVELVASPDRLLECTRYLGKNTFRTTYVDCAHLEANGPEOYVLSFDSRSGMGASHLSLT 600
 DB 550 atvkvsgvdgrldcettlgnktrfsvdgavletnqperlmnlfsdasgumaagpfsilt 609
 QY 601 YELPAGLOVYRISNGDCATPPPGGAPSAPEVNAFSCALRYRNFQORHSLTGLGM 660
 DB 610 yaasaaglevryvaagldhnavtagvsvrsapevnaafscalyrlnreaagvafslgnlw 669
 QY 661 LHPGGLGIPPEPSPGHWESANPFCGEGTLYTWTWS-TSGFSSDGFSP-----PEAAPA 714
 DB 670 fhpgglglipfsgpghwesanpfcggestlyrttsevdavspatpdlgfmsepsips 729
 QY 715 MAATPGLPBSTPYSDIWLVPPESEFQVDAAYV-PAAPPAGLP-GRVVLTPPPPPVH 772
 DB 730 raatppl-----aaplpapppppspsapalaepasgtag 766
 QY 773 KPSPR-PPSNRRLLYYPGAKVYAGSLPESODMWLVNNSNGHRRGGGICHAIFYGRF 831
 DB 767 apaltbqatnrrlltlypogskvlagstlscctcwlvnasnvdhrrpggdlchaifygrfp 826


```

Db 250 pIvIerayIghchIvIllaIteapepsmpIyVpYprsteyvrsIlgpgtIpsIlfptscstK 309
QY 301 STPIAVPVHIMDRIMLFGATLDDOAFCCSRIMLTURLGISYVWPGALVANGMNASSEDAL 360
Db -310 stIInavpnhIwIdrImIlgatIdDgaIccsIrmIcIlyIrgIsYvIvIcIvIvnaeqmaseaI 369
QY 361 TAXITAAVLTICHOXYLNTQAIKSGMRRLGVEHAKFTIRLISWLFKSGNDYIPGRLO 420
Db 370 taviIaayIltchqgrIylrtqIsgkmrIerehaqkIltIswlIfeksgdyIpgqIe 429
QY 421 FYACCRMLASGFHLDRVLVFDSSVPCRCFTFLKKAAGKCCMRMLGDECTGLEPAE 480
Db 430 fyaqcrIvIlaagfhlIdrvIvIldesapchcItrIkaVskCcctfmkVlgseCtflIpaE 489
QY 481 GLVDDHCHDNEAYGSEVDPEAPHLNVSGTYAVHGHOLEALYALANVPODIAARSLT 540
Db 490 gVvgdgqhdneaygsdVdpesaIsdIsayvVpIcalqIlyalIdIpaElvIaragrlt 549
QY 541 ATVELVASPRLCEKRTVLGNKTFRTVVDGAHLEANGPEQVLSFSDASROSMGASHLT 600
Db 550 atVvsgvdgrIdceIlgkIltfIstfVdgavIetngperhnlIsIdasqumaagIstI 609
QY 601 YELTPALQVRISNGIDCATFPPGAPSAAPGEVAAFGSALRYRNFTQRHSLTGLW 660
Db 610 yaasaagIevryaaqIdhnavIagIvysIrsapgevalafsalYIlnreagrlstIqnf 669
QY 661 LHPGLGIFPPSPGHIWESANPFCGEGTLRTWS-----TSGSSDSFSPPE 709
Db 670 fhpegIlgIrlafIepgIvhesaapIgcseIlylttsevdvpsapqdlIgfIs-----e 724
QY 710 AARPAAMATPEGLPHSTPPVSDIWLPPRSEFQVDAAPV-PPADPA-GLPGVYLTPRP 767
Db 725 pslpsraatp-----tp-----aaIpppdpdpIrlsapprgpap 761
QY 768 PPRVHKRSIP-PPSRNRRLIYTPDGAKYVAGSLFESDCMLVNASNPHRGGLCHAF 826
Db 762 gataaraIltqIarhrrIltfIypdsgkIvIagsIstscIvIvnaSvndhIpggIchaI 821
QY 827 YOREPEAFYPTFETIMREGLAAYTLTPRIIHAVAPDYRVEONPKRLAAAYRETCSRAGTA 886
Db 822 YqxyrasIdaasfImrIdgaayIltIprIihavapdyrIelnprIleaayIretcsrIgtA 881
QY 887 AYPLGSGITQVPVSLFSFDAMERNHNPGLDELXLEPANAPEAKPAQPVLTITEDTART 946
Db 882 ayplIlgIyIqvIpgIsIdawerIhIrgpIdelyIlpelaarIfeaarIpcIltIledvart 941
QY 947 ANLALETIDATEVGRACAGCTISPGIVHYOFTAGVPGSGKSRSIQGDVDVVVPPRELR 1006
Db 942 anIaIeldsaIdaIvgracaagIvIcpIvvyqIftagvpgsIksrIstIgaIvavvvppIreItr 1001
QY 1007 NSMRRRGFAAFTPTTAARVYIGRRVVIDEASLPPHLLLMORASSVHLIGDPNQIPAI 1066
Db 1002 nawrrIgfaaftptIaaIvgrIvIvIdeaPsIpphIllIhmgaIatvhlIlgdnpqIpaI 1061
QY 1067 DEFHAGIYVPAIREFLATSMWXYTHRCPADVCELIRGATYKIQTTSVYLSLEWNEPAIG 1126
Db 1062 dfehaGIvpaIrefIatIsmwXyThrcpadvcelIrgaypmIdqIstIsvIstIswgepavg 1121
QY 1127 OKLVYTOAAKAANGATVHEAGATFETTTIATADARGIIOSSRAHATVALTRHEKC 1186
Db 1182 vIdIapqIltIevgIsdaIvnnIflaageIghqpsvIprIpnandavtlIaIafpsccIsa 1241
QY 1247 YHOLAELGRAPVAVVLPCELEGGGLYMQDELTVSSVLVFEELTDIVHCMAAPSQ 1306
Db 1242 fheIaaelgrIapvavvlpIcpeIlegIlyIlpqelItcIsvvtIftIdIvchmaapsq 1301
QY 1307 RKAVLSTLVGRYGRRTLYAAHSADVRESLARFPTIGPOATCCELYELVEAMVEKGD 1366

```

```

Db 1302 rkaVslIvIrgygrIklIynashsdvrsIarIflaIpaIvqvIttcelIyeIleaamvekqgd 1361
QY 1367 GSAVLELDLNCNRDVSRIITFPFOKXCNKFTGTETIAHGKVGOGISAMSKTFCALGCPWFRAI 1426
Db 1362 gsaVleIdIcIrdvsrIltfIqkdcnkIlttgeIangKvvggIsawskIftcalIgpwItrai 1421
QY 1427 EKEIALLPPIIFGYDAYEESVFPAAYSGASGVWFENDESEPDSTONNPSLGIECYVME 1486
Db 1422 ekalIallpIpgvfygdIdtIvtsaavaaKaSmvfenIdseIstqnlIsIglecIame 1481
QY 1487 ECGMPQHLIRLHLVRSAMLIQAPKESLKGFWKHHSEPGTLIMNYWNAIIAHCEFR 1546
Db 1482 ecgmpqwlIrlIhIlsawIlgapkesIrgIvfkKhsgepIltIwIvWmaavIthcydIr 1541
QY 1547 DFRVAAFKGDSDVYLCDYOSRMAALLINGCGLKLVDRPGLYGVVAAVGLTLPD 1606
Db 1542 dIryaaIkkgdssIvIcseyIrgpIgaavIlgcgIklIvdrIpgIlyagvvaPgIgaIpd 1601
QY 1607 VVRPAGRLSEKMGPGPERAEQURLAVCDFLRGLTNNVQVCVDVSRVYGVSPGLVHNL 1666
Db 1602 vvrfagrlIteknvgpIperaeqIrlavsdflrklItnvaqmcvdvrsIvYgvpIghnll 1661
QY 1667 GMQITADGKAHFTETIKPVLDTNFIORVE 1698
Db 1662 emIqavagdgkahtfesvKpVIdItlnsIltcrve 1693

```

```

RESULT      8
W76368
ID W76368 standard; Protein; 1693 AA.
AC W76368;
XX
XX
DT 03-DEC-1998 (first entry)
DE Hepatitis E virus
KW Hollow particle protein; virus; antibody; detection; immunoassay;
KW infection.
OS Hepatitis virus.
XX
XX
FH Key Location/Qualifiers
FT Protein 1..1693
FT /note="Partial sequence"
XX
XX
PN JP10234383-A.
PD 08-SEP-1998.
PE 28-FEB-1997; 97JP-0062445.
PF 28-FEB-1997; 97JP-0062445.
PR 28-FEB-1997; 97JP-0062445.
XX
XX
PA (DENK-) DENKA SEIKEN KK.
PA (KOKU-) KOKURITSU YORO ETSEI KENKYUSHO.
DR N-PSDB; V61687.
XX
XX
PT Hepatitis E virus hollow particle poly(peptide(s) and nucleic acids
PT encoding it - useful for more accurate detection of HEV in samples,
PT using immuno-assays and nucleic acid hybridisation
XX
XX
XX Claim 10: Page 17-24; 29pp; Japanese.
XX
XX This sequence represents a Hepatitis E viral hollow particle protein.
XX This polypeptides can be used to raise antibodies to detect HEV
XX infection in samples, e.g. by immuno-assay based techniques, and the
XX nucleic acid can be used for the same in nucleic acid hybridisation
XX assays. The polypeptides and nucleic acids allow more accurate
XX detection of HEV than previously possible.
XX

```

SQ Sequence 1693 AA;

Query Match 82.5%; Score 7437; DB 19; Length 1693;
 Best Local Similarity 80.9%; Pred. No. 0;
 Matches 1391; Conservative 114; Mismatches 158; Indels 56; Gaps 5;

```

QY 1 PGTTTAEQAALAAANSALANAVVRPLSRVOTELLINLMOPROLVFRVLMNPIQR 60
   |||||
DB 10 pgltaeqaalaanaaanaavvrplshqgjeilllmoprgqlvfrvwmnpidqr 69
   |||||

QY 61 VHNLEJOYCARAGRCLEVGAPHSINPNVLAHRCFLRPVGRDQWRWTSAPTRGPAAN 120
   |||||
DB 70 vhnlelelycratsgrclelghaprsindpnvwhcflirpygrdqgrywrlaplrpaan 129
   |||||

QY 121 CRRSALRGLPPADRTYCFDGFSCAPAEETVALYSLHMLPADVAEAMARHGXTLYAA 180
   |||||
DB 130 crrsalrqlpavdrlycldqfsgcnfpaecgialylhdmspdsdaeamrthgmtrlyaa 189
   |||||

QY 181 LHPREVLPPGTYHTSYLLIHDGRVAVTYEGDSAGCNHDVSTLRAMRTKTIVGDH 240
   |||||
DB 190 lhppevllppgtyrtaasyllindgrvvlyegdsagynhdvsnlrswrltkvlgdh 249
   |||||

QY 241 PLVIERVRAIGHFVLLTAAPESPMPYVPYRSTEEVYRSIFGGSFSLPPSACSTK 300
   |||||
DB 250 plvieretralgchfvllltaapepsmpyvpypirsteeyrslfpgpgtprslptscstk 309
   |||||

QY 301 STEHAVVHIWDRMLRGATLDDQAFCCSHLMTYLRGISYKVTGVALVANEGRNASEDAL 360
   |||||
DB 310 sthnavahiwdrmlrgatlddqaecscrlmtylrgisykvtvgtlvaneegnasedal 369
   |||||

QY 361 TATITAAVLTICHQRLYLRQAISKGRRLGVENAKOFTITLYSWLREKSSRDVTPGQLO 420
   |||||
DB 370 taitaavlytichqrllyrqaaiskgmrllerehaqkflrplyswlreksrddytpgqle 429
   |||||

QY 421 FYAQCRRLMSAGFHLDPRLVLFDESVPRCRTEFLKIVAGFCFCFMKVLGCECFLPEAP 480
   |||||
DB 430 fyqcrtrmsagfhlprlvlfdesaprcrtelrkalskfcfcfmkvlgceccflpgae 489
   |||||

QY 481 GLVGDHGHNEAVEGSEVDPDAEPALHDVSGTYAVHGHOLEALYRALNVPDIDARASRLT 540
   |||||
DB 490 glvgdghghneavegsevdpaepalhdvsgtyavhgholealylralnvpdidaarasrlt 549
   |||||

QY 541 ATVELVASPRLRCRYVLTGKTRTYVVDGAHLEANGPEQYVLSFDSRDSMGASISLT 600
   |||||
DB 550 atvkvsvdgrldceellgnktrtsfvdgavleangperynlsfidsqstmaagrsfpt 609
   |||||

QY 601 YELTPAGLQVRISNGLDCTATFPFGAPSAPEVAAFCSALYRYRFRQHSITMGILW 660
   |||||
DB 610 yaasaaglevryvaagldhnavfapvyspsapgevtatfcsalylfrtreagrnslgnlw 669
   |||||

QY 661 LHPEGLGIRPPSPGHIMESANPFCEGTLYRTWS-TSGFSSDSP-----PEAAPA 714
   |||||
DB 670 fhpegllglrapsphwesanpfcegestlyrtwsevdavseparpdlglimseprlps 729
   |||||

QY 715 MAATPGLPSTPPVSDIWLPPPESEFQVDAAVPPAPDAGLPBVVLTTPPEPPYHNR 774
   |||||
DB 730 raatpvl-----aaplp-----lapdpppppsap 754
   |||||

QY 775 STIPP-----SRNRLLYTPDGAKYVAGSLFESDCDMLNANSGPHRG 819
   |||||
DB 755 aldepasgatagapaltghatarrllftypgdskvtagstfesaacvlnasvndltpg 814
   |||||

QY 820 GGLCHAFYORFPEAFYPTTEIMREGLAAYTLTPRPIIHAAVPDYRVADONKRLAAVRET 879
   |||||
DB 815 gglchafyypasfdaasftvmrdgaaycltprpilhavapyrlenhpnkrlleaayret 874
   |||||

QY 880 CSRGRTAAYVLLSGIYQVVSLSFDAMERNHRRGDELLYTEPAANKFEAKKPAQVLLT 939
   |||||
DB 875 csrlgtaayvllgtlyqvdlpsfdawernhrrpgdelylpelaarvfeanrprlclt 934
   |||||

QY 940 TETPARTANALFEDATEVGRACAGCTISPGIVHYOFTAGVGGSGRSRISOGGDVYVV 999
   |||||

```

```

DB 935 tedvartanaleldsatsdvygracagcrvtpyvgqftagvpgsgksrsitqadvvv 994
QY 1000 VPTRELNSKRRRGFAATPHTAARTIGRRVVIDAPSPRLPLLHMQRASVYLLGD 1059
   |||||
DB 995 vptrelnswrrrgfaatfphtraarvtgrrrvvdeapsrlpllllmqratvhlgd 1054
   |||||

QY 1060 PNOIPAIDEFHAGLVPAIRPELAPTSMMWYTHRCPCADVCELIAGVAPKIOTTSRVLSRF 1119
   |||||
DB 1055 pnoipaidfethaglvpalrpeplaptswwvthrcpadvcelttagvpmjutsrvlsrf 1114
   |||||

QY 1120 WNEPAIGOKLVYTOAKAANPAGATVNEAGATFETTTIATADAGLIQSSRAHAVAL 1179
   |||||
DB 1115 wnepagvqkrlvtfgaakaanpagsvvtneagatylcttliatadargllqssrahalval 1174
   |||||

QY 1180 TRHTEKCVILDAPGLIRENGISDYIVNFFLAGGEVGHRSYIPRGNPQDNGITQARP 1239
   |||||
DB 1175 trhtecksvildagllirengisdaivnfflaggeighrpsvrlpgrpndanvdtlaatp 1234
   |||||

QY 1240 PSCQISAYHQLAELGHRPAPVAAPLPPCELEOGLLYMPQELTVSDYLVFELTDIVHC 1299
   |||||
DB 1235 pscqisafhqlaelghrpapvaavlppcelevoglllympqelacdsyvtfeltdvnc 1294
   |||||

QY 1300 RMAAPSOAKAVLSTLVGRYGRRTKLYEAHSDVRESLARIPTIGVQATTCELYELVEA 1359
   |||||
DB 1295 rmaapgrkavlstlvgygrtcklynashdvreslarfipalpgvqvtcelyelvea 1354
   |||||

QY 1360 MWEKGODGSANVLELDICNRVSRITTFQKCKNFTTGTTIARHKVGOGISAMSKTCALP 1419
   |||||
DB 1355 mwekgodgsanvleldicndvrsriltfqqcknfttgetlahgkvqgisawaktcalt 1414
   |||||

QY 1420 GPWFRAIEREITLLLPNIFGYDAVESEVFAAIVSGSCGMENPFSEPDSTQNNPSTLG 1479
   |||||
DB 1415 gpwfraieretlllpnifgydaidevtsaavaakaamvendsetdsqnnfslg 1474
   |||||

QY 1480 LECVYNEECGMPQMLIRLYHLVRSAMTIOAPKESLKGFMKHSGEPTLLMNTWMNAIT 1539
   |||||
DB 1475 lecvyneecgmpqmlirlyhlrsawllqapkeslsgfwkhsgepgtllmnlwmnav 1534
   |||||

QY 1540 AHCTYERDFRVAAPFKGDDSVYLCSDYRQSRNAALTAGCGLKLVDRPGLYAGVYAP 1599
   |||||
DB 1535 ahctydrfdlqvaaftkgddsvlylcseyrqpagaavllagcglklkvdrpilyagvvyap 1594
   |||||

QY 1600 GLGTLPVYRFAGRLSKNGMGPBERAEOLRLVCDPLRLTNVAOVCVVYRNVGVSP 1659
   |||||
DB 1595 glgtlpvvrfaagrlskngmgsrperadelrlvdsdflrltnvaqmcvavsvrygvsp 1654
   |||||

QY 1660 GLVHNLIGMLQTLADGKAHFTETIKPVLDTNSIIOHVE 1698
   |||||
DB 1655 glvhnligmlqavdgakahfteesvkvldltnslcive 1693
   |||||

RESULT 9
W71209
ID W71209 standard; Protein; 1693 AA.
XX
AC W71209;
XX
DT 30-OCT-1998 (first entry)
XX
DE Protein encoded by ORF 1 of the Burmese isolate of ET-NANB.
XX
KW Enterically transmitted nona/nomb hepatitis virus; identification;
KW HEV; ET-NANB; detection; vaccine.
XX
OS Hepatitis virus.
XX
FH Key Location/Qualifiers
FT MISC-difference 154 /note= "not specified"
FT MISC-difference 1514 /note= "not specified"
FT MISC-difference 1552 /note= "not specified"
FT MISC-difference 1552 /note= "not specified"

```

```

XX  US5789559-A.
PN
XX
PD  04-AUG-1998.
XX
PF  25-JUN-1994; 94US-0279823.
XX
PR  05-APR-1991; 91US-0681078.
PR  17-JUN-1988; 88US-0208997.
PR  11-APR-1989; 89US-033672.
PR  19-JUN-1989; 89US-0367486.
PR  13-OCT-1989; 89US-0420921.
PR  05-APR-1990; 90US-0505888.
PR  25-JUL-1994; 94US-0279823.
XX
PA  (GENE-) GENELABS TECHNOLOGIES INC.
PI  Bradley DM, Fry KE, Krawczynski KZ, Reyes GR, Tam A;
PI  Yarbough FO;
XX
DR  WPI: 1998-446186/38.
DR  N-PSDB: V54729.
XX
PT  Hepatitis E virus DNA - useful for e.g. virus detection and viral
PT  protein production
PS  Disclosure: Columns 55-64; 45pp; English.
XX
XX  W71209-11 represent the proteins encoded by the open reading frames
CC  (ORFs) of the DNA sequence of the Burmese isolate of an enterically
CC  transmitted nonA/nonB viral hepatitis agent (ET-NANB). The nucleic
CC  acid sequence may be used for identifying and sequencing the entire
CC  viral agent (also referred to as HEV), detecting ET-NANB in
CC  infected samples, e.g. by specific amplification of virus-derived DNA
CC  sequences and for producing recombinant viral proteins for use in
CC  vaccines.
XX
XX  Sequence 1693 AA;
SQ

```

Query Match 82.5%; Score 7437; DB 19; Length 1693;
 Best Local Similarity 81.7%; Pred. No. 0;
 Matches 1394; Conservative 110; Mismatches 171; Indels 32; Gaps 6;

```

QY  1 PGITTAISOALIAANSAALANAVVVRPFLSRVQTEILINLMQPRQVFRPEVIANHPRIOR 60
DB  10 pgitaisqaaalaanaasalanavvrflshqieilnlmqprqlvtrpevfahnpqtr 69
QY  61 VIHNELOYCARAGRCLEVGAAHPSINDPNVLRHCFLRPVGRDVQRWYSAPTRGPAAN 120
DB  70 vihnelycraargrcleigahpsindpnvhrclfrpvgdvqrwyatrgpaaan 129
QY  121 CRRSALRLPRADRTYCCDFSRCAFAETGVATLSLHDMRADVAEMARIGXRLTAA 180
DB  130 crsralrlpradrtycdfisrcfaetglatalslhdmrpsdvaeamifrgmlryaa 189
QY  181 LHLPEEVLPPGTYHTTSTYLLIHODGRAVVYEGPTSGAGYVNDVSLTAAWIRTKTIVDDH 240
DB  190 lhlpeevllppgtyhtstyyllihodgravvyegptsagynhdvsltawirtktkvtdh 249
QY  241 PLVIERVRAICHFVLLTAAPEPSMPYVPRSTEVYVRSIFPGSGPSLIFSACSTRK 300
DB  250 plviervraicHFVLLTAAPEPSMPYVPRSTEVYVRSIFPGSGPSLIFSACSTRK 309
QY  301 STEFANVPVHMDRLMFCATLDDOAFCCSKMTYIRGISTYVTVGALVANEGWNASSEDAL 360
DB  310 stefanvpvhmdrlmfcatlddoafccskmtyrigistvytvvgalvanegwnasedal 369
QY  361 TAXITAAVLTICHOXYLTOAISGMRRLGVEHAKOFTIRLXSMLEESGSDYIPGRLO 420
DB  370 taxitaaVLTICHOXYLTOAISGMRRLGVEHAKOFTIRLXSMLEESGSDYIPGRLO 429
QY  421 FYAOCRWLSAGFHLDPRLVLFDESVPICRRTFLKVKVAGKFCCEFRWMLGOCCTCLEPAE 480

```

```

DB  430 fyaocrwlsagfhlDPRVLVLFDESVPICRRTFLKVKVAGKFCCEFRWMLGOCCTCLEPAE 489
QY  481 GLVGHGHDNENAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNYPQIARASRLT 540
DB  490 glvghghdnENAYEGSEVDPAEPALHDVSGTYAVHGHOLEALYRALNYPQIARASRLT 549
QY  541 ATVELVASPDRLCEKRTVVGKNTFRTTVDGALHLEANGSEGVLSFSDASROSGAGSHSLT 600
DB  550 atvELVASPDRLCEKRTVVGKNTFRTTVDGALHLEANGSEGVLSFSDASROSGAGSHSLT 609
QY  601 YELPPAGLVAVRISNGLDCTATFPFGAPSAAPGEVAAFCSALYRYNRTQRHSLTGLG 660
DB  610 yELPPAGLVAVRISNGLDCTATFPFGAPSAAPGEVAAFCSALYRYNRTQRHSLTGLG 669
QY  661 LHPEDLLGIFPPFSRGHFWESANPRCSGCTLYTRKWS-TSGFSSDFSP-----PEAAPA 714
DB  670 lhpEDLLGIFPPFSRGHFWESANPRCSGCTLYTRKWS-TSGFSSDFSP-----PEAAPA 729
QY  715 MAATPGLPHSTPPVSDIWLPPPSSEFQVDAAPV-PPAPDPAGLP-GPVVLTpppppEVH 772
DB  730 raaptl-----aaPlpppdpdpppsapaIeapagatag 766
QY  773 KPSIP-PPSRNRRLTYTPDGAKYVAGSLFESDCDMLVNASNPGHRRPGGLCHAFYGRFP 831
DB  767 apaltbqtrahrlllftypdgskvfagslfestctwlvnasvndhrpggylchafyqtryp 826
QY  832 EAFYPTFIMEEGLAAYTLTPRIIHAVAPDYRVQNFKLEAAYRETCSRRGTAAVPL 891
DB  837 asfaasfvmddgaagayllcprlilhavapyrllhmpkrlleaayretcsrlgaaypll 886
QY  892 GSGIYOVVSLSPFQAMENHRPGDELTYTEPAAMFEENKRAQVOLTTEGTARTANL 951
DB  887 gtgilyqvpigsfdaewnhnpgdelylpelaaaayfeenrtprrlltledvatantlai 946
QY  952 EIDATEVGRACAGCTSPGIVHYQFTAGVPGSKRSIQOGDVVVVPPRELNSMR 1011
DB  947 eldatdvgracagcgvryvvyqftagvpgsgkrsitgadavvvpprelnawtr 1006
QY  1012 RGFAPFPTHTAARTTIGRRVVIDEAPSLPHLLLLHMQRASVHLGDPNQIPADIEHA 1071
DB  1007 rgaafphtaarttigrvvvidaeapslphllllhmqraatvhlldgpnqipaideha 1066
QY  1072 GLVPAIRPELAPTSMWXXVTHRCPADVCELRGAPVKIOTTGRVRSLEFNPALAGOKLVX 1131
DB  1067 glvpaiprdlptsmwvthrcpadvcelrgapvmqitgrvrslefngpavagqrlvf 1126
QY  1132 TQAAKANPGAITYHEAGATFTETIIATADANGLIQSSRAHAIVALTRHTEKCVILDA 1191
DB  1127 tqaakpangavstvehaqatftetliatadargllqssrahaivaltrhtekcvilda 1186
QY  1192 PGLIRENGISDVYVNNPFLAGGEVGHRRPSYIPRGNPDONLGTIQAPPPSQOIAHYOLA 1251
DB  1187 pglirevgisdaivnnflaggeighrpsviprgpnpanvdltaafppscqiaafibqla 1246
QY  1252 BELGHRPAPVAAVLPPCELEOGLLYMPQELTVSDSVLPEFLTQIVRCMAAPGQRAVL 1311
DB  1247 belghrpyvaavlppeleoglllympqeltdcsavvleltdiivncmaapqgraxvl 1306
QY  1312 STLVGRYGRRTKLYEAHSDVRESIARPTIIPGVOATTCELYELVEMAVEKGGDGSAYL 1371
DB  1307 stlvgyrgrrtklynaasdvreslarfpaiipyvqttecllyelvaamekvgdgsayl 1366
QY  1372 ELIDCNRDVSRTTFQCKXNKFTTGETIAGKVQGGISAMSKTFICALGFWRAIEKEIL 1431
DB  1367 elidcnrdsrttfqcknkfttgetiagkvqggisawsktficalgfwfraiekaill 1426
QY  1432 ALLEPNIFYGDAVEEYFAAASGASGCMWFENFSDSTONNFSGLCGYVWEEGMP 1491
DB  1427 allepnifygdavdeeyfaaasgasmwfendfseidstqnfnfsglcoalmeeegmp 1486
QY  1492 QMLTRLHLVRSAAVLIQAPESLKGFWKHSGEGETLLMNTVMMAITACYFERDRVA 1551
DB  1492 qmltrlhlvrsAAVLIQAPESLKGFWKHSGEGETLLMNTVMMAITACYFERDRVA 1551

```

Db 1487 gwllrlyhlrsawllgqpkesslrgfwkxhsgepgllmwlvmmavithcydfdfgva 1546
Qy 1552 AFKGDSDVVLSDRQSNAAALLTAGCGLKLVKVRPGLGAGVAVAGLSTLPPVAF 1611
Db 1547 afkgdxsivlseyrqspsaavllagcgklkvdfirpigliyagvavapglapdvrf 1606
Qy 1612 GRLSKNNPGPERAEOLRLVCDPLRGLTNVAVCVVSVRVGVSGLVHNLIGMLQ 1671
Db 1607 grlckwpgpberaeqrlrlavsdllrlkltnvaqmcvavsrlygvsplvnhllgmld 1666
Qy 1672 IADKAEHETIKPYLDLJNSIQRVE 1698
Db 1667 vadgkafhtesvkpyldltnsilcrve 1693
RESULT 10
R51264 ID R51264 standard; Protein; 1693 AA.
XX AC R51264;
XX DT 21-OCT-1994 (first entry)
XX DE HEV strain protein encoded by ORF-1.
XX KW Hepatitis E virus; HEV; strain SAR-55; open reading frame; ORF;
XX KM antibody; detection; diagnosis; primates; stool suspension.
XX OS Hepatitis E virus strain SAR-55.
XX PN WO9406913-A.
XX PD 31-MAR-1994.
XX PF 17-SEP-1993; 93WO-US08849.
XX PR 18-SEP-1992; 92US-0947263.
XX PA (USSH) US SEC DEPT HEALTH.
XX PI Emerson SU, Purcell RH, Tsarev SA;
XX DR WPI: 1994-118462/14.
XX DR N-PSDB: Q45197.
XX PT Purified hepatitis E strain SAR-55 virus - used to develop prods.
XX PT for use in detection, diagnosis, vaccines and therapy of
XX PT hepatitis E virus infection
XX PS Disclosure; Page 57-62; 114pp; English.
XX CC The sequences given in R51264-66 are encoded by the hepatitis E virus
XX CC (HEV) strain SAR-55. The CDNA sequence contains three open reading
XX CC frames (ORFs). These proteins can be used to stimulate the production
XX CC of protective antibodies upon injection into a mammal that would serve
XX CC to protect the mammal upon challenge with wild type HEV. The proteins
XX CC can be used for detection and diagnosis of HEV infection. The HEV
XX CC SAR-55 CDNA was isolated from primates inoculated with stool
XX CC suspensions obtained from hepatitis E patients.
XX SQ Sequence 1693 AA;
Query Match 82.5%; Score 7436; DB 15; Length 1693;
Best Local Similarity 81.5%; Pred. No. 0;
Matches 1396; Conservative 112; Mismatches 162; Indels 42; Gaps 7;

Db 70 vihnlelycrasgrclelgahpsindnpvvhrcflrpadrdvgrvltaptrgpaan 129
Qy 121 CRRSALRLPLPADRTYCGDGSRCAPFAETGVALLYSLDMLPADYAEAMARHGXRLLAA 180
Db 130 crrsalrlglpadrtycfdgtsrcnfpaelajalslmdpsdvaeamfthgmkrlyaa 189
Qy 181 LHLPEVLLPGTYHTTYSYLLIHODGRAVVTYEGGTSAGYHNHDSILKAWIRTPRTYVDH 240
Db 190 lhlpevllppgyrftaaylllhdgrvrvlyegdtlsagynhdvsnlslrtlkvtggh 249
Qy 241 PLVIERVAICHVYLLTLTAAPESPMPYVYPRSTEVYVRSIFEGGSPSLFSPACSTK 300
Db 250 plviervraichvfyllltaapepmpyprstevyvrslfeggsptslfpscstlk 309
Qy 301 STEFANVPVHMDRLMFCATLDDAFCGSRMLTYLRGISTKYVTGALVANSNMASDAL 360
Db 310 stfnavpahldwrlmflgatliddgafccsrilmtylrgisvrvtyglvanegmasedal 369
Qy 361 TAXITAAVLTICHOHRYLRTQAIKSGMRRLGYEHAGKFTIRLYSMLFEKSGRDIYGRLO 420
Db 370 taxitaaylvlichnqrylrltqalskymrlterhaqklrlrlyswlfeksgrdiyprgqe 429
Qy 421 FYAQCRRLISAGFHLDPVLFVDESVPCRCRTFLKRVAGKFCCEMRMLGOECTCLEBAE 480
Db 430 fyaqcrtrlasgfhldprvlyfdesapchcrtalrkavskfcfcmkwlqgectclgpa 489
Qy 481 GLVGDHGDNEAYEGSEVDPAEPALHDVSGTYAVHGHOLEALRYALNPQDIARASRLT 540
Db 490 gvgdgqhdneayegsdvdpaesaidsisgyvvgtalqpygaltdipaelvaragrlt 549
Qy 541 ATVELVASPDRLRECTVYNGNTFRFTYVDGAHLENGPEQYVLSFSDARSOSGASHSLT 600
Db 550 atvkvsvqygrldccelllgnkltfcsfydgavlecnperhmlsdaasqstaapgfsl 609
Qy 601 YELTPAGLOVRISNGLDCTATFPFGCAPSAAPCEVAFCALRYRNFTQSRHSLTGGLM 660
Db 610 yaasaaglevryaaagldhravfapgvpsrpagvetafcslaylfnreaqrllstgnfw 669
Qy 661 LHPEGLIGIFPPFSFGHIMESANPPCGEGTLYTRTWS-----TSGFSDFSPPE 709
Db 670 fhpegllgfpafspghwesanpfcgscltyrltwevadavspagdlgfts-----e 724
Qy 710 AAAPMAATPGLPHSTPPVSDIWLPPSEEFQVNAAY-PPAPPA-GLPEPVVLTLP 767
Db 725 psiperaatp-----lp-----aaipppadpsptlspargepap 761
Qy 768 PPPVHKPSIP-PPSRNRLTYTPDGAKYAGSLFESDCDMLVNASNPGHRGGGLCHAF 826
Db 762 gatarapatlqtatrrlllftypdgskvfagslfestcvtlvnasnvdhprggllchaf 821
Qy 827 YORPEAFYPTFEIMREGIAVYTLPRPIIHAADVREONPKRLDAVRETCSRRTGA 886
Db 828 ygrypasidaafvnrdaaaytlcprllhavadyrlehnpkrleaaayretcsrlgta 881
Qy 887 AYPILGSGIYQVPLSLPDAMERNHRPGDELTYLPANMWFANPAPQVPLTTEDTAT 946
Db 882 aypllgltqlyqvpilpsldawernhrpdelylpelaawfleanpccpctlltcaavrt 941
Qy 947 ANLALEIDAATEVGRACAGCTISPGIVHYQTFAGVPGSGKRSIOGDVDDVVVYPTREL 1006
Db 942 anlaleidsatcdvgracagcrrtppgvqyqftagvpgsgkrsrlsgaadvvvptrelr 1001
Qy 1007 NSMRRRGFAALTPTHTAARTYIGRRVYIDAEPLPHLLILHQRASSVHLCDPNOIPAI 1066
Db 1002 nsmrrrgfaaltpthtaartygrrvlydeapslphllllhmqraetvhllygdpnqipa 1061
Qy 1067 DEHAGIVPAIRPELAPISMWXVYTHRCADYVCELIRGAVPKIOTQTSRYLRSIFMNEPAT 1126
Db 1062 dfehagivpaairpdlaplswwvthrcpadvcelltrgavpmqltsrslrslfswgepav 1121
Qy 1127 OKLVYTOAKKANPAIIVVHEOGATFETTTIATADAKGLIOSSRAHAIVALTHTTEKC 1186
Db 1122 qklvftgaakaanpsvtrcheagatyetclltatadargllqssrahaivaltthtekc 1181

QY 1187 VILADAGLLREVGISDVYNNFLAGGEVGHRRPSVIRPCNPONDLTQAPPSQISA 1246
 DB 1182 VILADAGLLREVGISDAVNNFLLAGGEVGHRRPSVIRPCNPONDLTQAPPSQISA 1241
 QY 1247 YHOLAEELGRAPVAAVLPCELEGLLYMQELTSSVSVFELTIDVHCRMAAPSO 1306
 DB 1242 YHOLAEELGRAPVAAVLPCELEGLLYMQELTSSVSVFELTIDVHCRMAAPSO 1301
 QY 1307 RKAVLSTLVGRYGRRTKLYAAHSADVRESLARTPTIGPQATTCETCELYELVEMVEKGO 1366
 DB 1302 RKAVLSTLVGRYGRRTKLYAAHSADVRESLARTPTIGPQATTCETCELYELVEMVEKGO 1361
 QY 1367 GSAVLELDLCNRDVSRTTFQKXNKFTTGETTAGKVGSGISAMSKTCALFGPMFRAT 1426
 DB 1362 GSAVLELDLCNRDVSRTTFQKXNKFTTGETTAGKVGSGISAMSKTCALFGPMFRAT 1421
 QY 1427 EKRTLLLPNTFYGDVAVESVFAAASGSGMVFENPSESDSQNNNSLCEVME 1486
 DB 1422 EKRTLLLPNTFYGDVAVESVFAAASGSGMVFENPSESDSQNNNSLCEVME 1481
 QY 1487 ECGMPOMLIRLYHLVSVAMTLQAPKESLKGFMKHSGEPTLLMNTVMNAITAHCEFR 1546
 DB 1482 ECGMPOMLIRLYHLVSVAMTLQAPKESLKGFMKHSGEPTLLMNTVMNAITAHCEFR 1541
 QY 1547 DEFVAAPKDDSVYLCSDYRQSRNAALAGCGLKLVDRPIGLYAGVVAPGCTLPD 1606
 DB 1542 DEFVAAPKDDSVYLCSDYRQSRNAALAGCGLKLVDRPIGLYAGVVAPGCTLPD 1601
 QY 1607 VYFAPARLSKMGNGPPEPEREQLAVCDRLGTNAQVCVAVSVRVSGLVHNL 1666
 DB 1602 VYFAPARLSKMGNGPPEPEREQLAVCDRLGTNAQVCVAVSVRVSGLVHNL 1661
 QY 1667 GMLQTTADGKAPHEFTIKPVLDTNSIIQRYE 1698
 DB 1662 GMLQTTADGKAPHEFTIKPVLDTNSIIQRYE 1693
 RESULT 11
 R14618
 ID R14618 standard; Protein; 1693 AA.
 XX
 AC R14618:
 DT 16-JAN-1992 (first entry)
 XX
 DE Protein encoded by ORF 1 of Burmese ET-NANB viral strain.
 KW enterically transmitted non-A, non-B hepatitis virus; hepatitis C;
 KW HCV; E.coli strain BR4; ATCC deposit number 67717; Burma.
 OS Enterically transmitted non-A, non-B hepatitis virus.
 PN MO9115603-A.
 XX
 PD 17-OCT-1991.
 XX
 PE 05-APR-1991; 91MO-US02368.
 XX
 PR 05-APR-1990; 90US-0505888.
 XX
 PA (GENE-) GENELABS INC.
 PA (USSH) US DEPT HEALTH & HUMAN.
 XX
 PI Reyes GR, Yarbough PO, Bradley DW, Krawczynski KZ, Tam A;
 PI Fry KE;
 DR WPI; 1991-325242/44.
 DR N-PSDB; Q14412.
 PT New viral proteins from non A-non-B hepatitis agent - used to
 PT treat and prevent enterically-transmitted non-A non-B hepatitis
 PT virus

XX
 PS Disclosure; Page 15; 117pp; English.
 CC A positive clone ETL1 was identified in a library prepared from
 CC bile of cynomolgus monkeys infected with the Burma strain of ET-NANB.
 CC Both strands of ETL1 were sequenced. Identity of the sequence with
 CC sequences in etiologic agents has been confirmed by locating a
 CC similar sequence in a viral strain isolated in Burma. This protein
 CC is encoded by the longest ORF (ORF 1) of the Burma strain.
 CC (See Q14410 for ETL1).
 XX
 SQ Sequence 1693 AA:
 Query Match 82.1%; Score 7400; DB 12; Length 1693;
 Best local Similarity 81.1%; Pred. No. 0;
 Matches 1388; Conservative 110; Mismatches 172; Indels 42; Gaps 6;
 QY 1 PGTTAIEQALAAANSALANAVVRPPLSRVOTETLLNMQPRLVFRPEVLMNPIOR 60
 DB 10 pgtlaieqalaanaalanavvrpplshqjeillnmqprqlvfrpevfmhpiqr 69
 QY 61 VHNELDYCRARAGCGLEVGAPRSINDPNVLRHCPRLPVGROVQRYSAPTRGPAAN 120
 DB 70 vhnelycraragcglevgaprsindpnvvrhrcplrpvgrovyqlaprtgpaan 129
 QY 121 CRSSALRGAPPADRYCFDGFSCAFAEFGVALYSLHDLMPADVAEAMARHGXTRLYAA 180
 DB 130 crssalrgappadrtcyldgfgscnfaetgialylhdmgspeaegmfrgmtrryaa 189
 QY 181 LHLPEVLLPFGYHTTSTYLLHGDRAVYTYEGDTSAGYNHDSILRAMIRTKLVGCH 240
 DB 190 lhlpevllpfgyhttsyllhgdraavytyegdtsagynhdsilrwmirktkvchg 249
 QY 241 PLVIEVRALGCHFFVLLTAAPPEPMPVYPRSRVEVRSIFGPGSPLEPSKSTK 300
 DB 250 plvieralghcfvlltaapepmpvyprrsrvevrsifpgspplpskstsk 309
 QY 301 STFHAAVPHIMRLMFGATLDQAFCCSRMLMTYLRGYSKYVGAALVANEGNASEDAL 360
 DB 310 stfhavphimrlmfgatldqafccsrmlmtylrgyskyvgaalvanegnasedal 369
 QY 361 TATITAAVLTICQRLRTRQVATSKGMRRLGVENAKOFTIRLSMFEKSGROYTPROQ 420
 DB 370 tavitaaylvlticqrlrtrqvatskgmrllgvnakoftirslsmfeksgroytproq 429
 QY 421 FYAQCRRLWSAGFHLPRLVLFDESVCORFLPKVAGCFCECFMLOCEGCFEPPE 480
 DB 430 fyagcrrlwsagfhlprrlvlfdesvcorflpkvagcfcecfmlocegcfeppae 489
 QY 481 GLVGDHGHNEAYEGSEVDPAEPANLDVSGTYAVHGHOLEALYRALNVODIAARASRLT 540
 DB 490 glvgdghghneayegsevdpaepanldvsgtyavhgholealyralnvdiaarasrlt 549
 QY 541 ATVEVVASPDRLCEKRTVLCNKTFRITTVVDAHLNAGPEQVYLSFPAOSKMGASHSLT 600
 DB 550 atvevvaspdrlcekrvlgcnktrfritvvdahlngapeqvylsfpaoskmgashslt 609
 QY 601 YELTPAGLOVRISNGLDCTATFPFGGAPSAAPGVAACSAALYRNRRTORSLTGLGL 660
 DB 610 yaasaglevryyaaqldhnavfapgfspasgvdalatsalyrnrearsltglgl 669
 QY 661 LHPEGLGIFPPSPGHIMESANPFCGEGTLTYRTWS-TSGFSSDPS-----PEAABA 714
 DB 670 fhpeglgifppspghimesanpfcgegtlyrtws-tsgfssdps-----peaaba 729
 QY 715 MAATPGLPHSTPPVSDIWLPPPEEFQYDAAPV-PPADPPAGLPGVVLTLP----- 766
 DB 730 raatppl-----aaipppapdppppsaaplaepasatgg 766
 QY 767 PPPPVAKRSTPPSRNRRLITTYPGCAKYAGSLPESDDMLVNASNPGRGSGGLCHAF 826
 DB 767 apaltbq-----tarnrtllfypdskvfaagstestctwlvnasnvhrp99glchaf 821

Thu May 31 09:15:05 2001

Search completed: May 30, 2001, 16:09:22
Job time: 199 sec

us-09-468-147-91.rag

Page 18

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 30, 2001, 16:09:23 ; Search time 70.93 Seconds

(without alignments)
2805.851 Million cell updates/sec

Title: US-09-468-147-91

Perfect score: 9013
Sequence: 1 PGTTTAEQAALAAANSALA.....FTETKPVLDLTNSIQRVE 1698

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: SPREMBL_15:*
2: sp.archaea:*
3: sp.bacteria:*
4: sp.fungi:*
5: sp.human:*
6: sp_invertebrate:*
7: sp.mammal:*
8: sp.mhc:*
9: sp.organelle:*
10: sp.phage:*
11: sp.plant:*
12: sp_rodent:*
13: sp.unclassified:*
14: sp.vertebrate:*
14: sp.virus:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	9001	99.9	1698	14	09YLK3
2	8812.5	97.8	1708	14	09YK10
3	8780.5	97.4	1708	14	09YK10
4	7730	85.8	1707	14	09YK10
5	7468	82.9	1693	14	09YK10
6	7457	82.7	1693	14	09YK10
7	7456	82.7	1693	14	09YK10
8	7424	82.4	1693	14	09YK10
9	7412	82.4	1693	14	09YK10
10	7333	81.4	1693	14	09YK10
11	7329	81.3	1693	14	09YK10
12	7304	81.0	1693	14	09YK10
13	7301	81.0	1693	14	09YK10
14	3444	38.2	727	14	081873
15	2520	28.0	479	14	071144
16	2365	26.2	452	14	071145
17	2284	25.3	487	14	081869
18	1821	20.2	491	14	09YK12
19	1368	15.2	290	14	081875

20	1183	13.1	245	14	081868	081868 hepatitis e
21	1081	12.0	225	14	09WKE5	09WKE5 hepatitis e
22	1074	11.9	217	14	081864	081864 hepatitis e
23	1025	11.4	212	14	09YK11	09YK11 hepatitis e
24	988	11.0	210	14	092745	092745 hepatitis e
25	988	11.0	210	14	092746	092746 hepatitis e
26	988	11.0	210	14	092747	092747 hepatitis e
27	988	11.0	210	14	092748	092748 hepatitis e
28	988	11.0	210	14	092749	092749 hepatitis e
29	981	10.9	210	14	056046	056046 hepatitis e
30	913	10.1	193	14	09WKE6	09WKE6 hepatitis e
31	884	9.8	182	14	081863	081863 hepatitis e
32	769	8.5	158	14	09W952	09W952 hepatitis e
33	763	8.5	152	14	081874	081874 hepatitis e
34	709	7.9	134	14	09YK14	09YK14 swine hepat
35	692	7.7	158	14	081867	081867 hepatitis e
36	666	7.4	136	14	09WKE3	09WKE3 hepatitis e
37	665	7.4	136	14	09WKE7	09WKE7 hepatitis e
38	658	7.3	136	14	09W9Y6	09W9Y6 hepatitis e
39	654	7.3	136	14	09WKE8	09WKE8 hepatitis e
40	646	7.2	136	14	09WKE9	09WKE9 hepatitis e
41	643	7.1	123	14	09QRR7	09QRR7 hepatitis e
42	639	7.1	123	14	09QRR6	09QRR6 hepatitis e
43	638	7.1	123	14	09QRR8	09QRR8 hepatitis e
44	618	6.9	160	14	081865	081865 hepatitis e
45	514	5.7	1704	14	067724	067724 helicoverpa

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	1698 AA.
09YLK3	09YLK3			
AC	09YLK3			
DT	01-MAY-1999 (TREMblrel. 10, Created)			
DT	01-MAY-1999 (TREMblrel. 10, Last sequence update)			
DT	01-JUN-2000 (TREMblrel. 14, Last annotation update)			
DE	POLYPROTEIN (FRAGMENT).			
GN	ORF1.			
OS	Hepatitis E virus.			
OC	Viruses: ssRNA positive-strand viruses, no DNA stage.			
OX	NCBI_TaxID=12461;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=HEV-US1;			
RX	MEDLINE=99190429; PubMed=10092008;			
RA	Erker J.C., Desai S.M., Schlauder G.G., Dawson G.J., Mushahwar I.K.;			
RA	"A hepatitis E virus variant from the United States: molecular			
RT	characterization and transmission in cynomolgus macaques.";			
RT	J. Gen. Virol. 80:681-690(1999).			
RL	[3]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=HEV-US1;			
RC	Schlauder G.G., Erker J.C., Dawson G.J., Desai S.M., Knigge M.F.,			
RA	Kwo P.Y., Smalley D.L., Rosenblatt J.E., Mushahwar I.K.;			
RA	Submitted (NOV-1997) to the EMBL/Genbank/DBD databases.			
RL	[4]			
RN	SEQUENCE FROM N.A.			
RP	STRAIN=HEV-US1;			
RC	Erker J.C., Schlauder G.G., Dawson G.J., Desai S.M., Mushahwar I.K.;			
RA	Submitted (APR-1998) to the EMBL/Genbank/DBD databases.			
RL	EMBL, AF060668; AAD15812.1;			
DR				

DR INTERPRO: IPR000606; -
 DR INTERPRO: IPR002588; -
 DR INTERPRO: IPR002589; -
 DR PRAM: PR01443; Viral_helicase1, 1.
 DR PRAM: PR01660; Vmehylitransf, 1.
 DR PRAM: PR01661; DUF27, 1.
 KW Polypotein.
 FT NON_TER
 SQ SEQUENCE 1698 AA; 186069 MW; 1C670601150F64EA CRC64;

Query Match 99.9%; Score 9001; DB 14; Length 1698;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1698; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PGITTAEOAALANANALANAVVVRPLSVOTETILINMOPQOLVRRPEVLNHHIOR 60
 DB 1 PGITTAEOAALANANALANAVVVRPLSVOTETILINMOPQOLVRRPEVLNHHIOR 60
 QY 61 VINHEBOYCRARAGRCLEVGAPHSINDNPNVLRHCFELRPVGRDVRWYSAPTRGPAN 120
 DB 61 VINHEBOYCRARAGRCLEVGAPHSINDNPNVLRHCFELRPVGRDVRWYSAPTRGPAN 120
 QY 121 CRNSALGFLPPADRTYCFDFGSCAFPAETGVALLYSLHDLMPADVAEMARHGXTRLYAA 180
 DB 121 CRNSALGFLPPADRTYCFDFGSCAFPAETGVALLYSLHDLMPADVAEMARHGXTRLYAA 180
 QY 181 LHPPEVLLPGGYHTHTSYLLIHGDRAVVYEGDTSAGYNHDSILRAMTRTKIYGDH 240
 DB 181 LHPPEVLLPGGYHTHTSYLLIHGDRAVVYEGDTSAGYNHDSILRAMTRTKIYGDH 240
 QY 241 PLVIERVRAIGCHEVLLTAAPEPSMPVYPRSTEVYVRSIFPGSGPSLFPSSACSTK 300
 DB 241 PLVIERVRAIGCHEVLLTAAPEPSMPVYPRSTEVYVRSIFPGSGPSLFPSSACSTK 300
 QY 301 STEHAVVHIMDRMLMELGATLDQAFCCSRMLTMRGISTKVMYGAALVANEWMASDAL 360
 DB 301 STEHAVVHIMDRMLMELGATLDQAFCCSRMLTMRGISTKVMYGAALVANEWMASDAL 360
 QY 361 TAXITAAVLTICHORYLDRTOAISKMRRLGVEHAKFITRLYSWLFPSKSGRDYIPGHOLO 420
 DB 361 TAXITAAVLTICHORYLDRTOAISKMRRLGVEHAKFITRLYSWLFPSKSGRDYIPGHOLO 420
 QY 421 FYACORRWLSAGFLHDPRLVLFDESVCRCRTELKKYAGKRCFCMRLGOBCTCFLEPAE 480
 DB 421 FYACORRWLSAGFLHDPRLVLFDESVCRCRTELKKYAGKRCFCMRLGOBCTCFLEPAE 480
 QY 481 GLVGDHNDNEAYEGSVDDAEPALHDVSGTYAVHGHOLEALYRALVNPDDIARASRLT 540
 DB 481 GLVGDHNDNEAYEGSVDDAEPALHDVSGTYAVHGHOLEALYRALVNPDDIARASRLT 540
 QY 541 ATVELVASPRLBECRTVLGNKTFRTYVDDAHLEANGPEOYVLSFDASROSMGASHSLT 600
 DB 541 ATVELVASPRLBECRTVLGNKTFRTYVDDAHLEANGPEOYVLSFDASROSMGASHSLT 600
 QY 601 YELTPAGLOVRISNGIDCATTPPGGAPSAAGEVAFFCSALYRYNRFTQRHSLTGLW 660
 DB 601 YELTPAGLOVRISNGIDCATTPPGGAPSAAGEVAFFCSALYRYNRFTQRHSLTGLW 660
 QY 661 LHPFGLIGITPPSPGHIWMSANPFCGEGTLTWTSTSGFSSPSPPEAAAPAMATPG 720
 DB 661 LHPFGLIGITPPSPGHIWMSANPFCGEGTLTWTSTSGFSSPSPPEAAAPAMATPG 720
 QY 721 LPHSTPVSIDWILPPSEEFQVDAAPVPAPAPAGLPGVVLTPPPPVHKSIPSPS 780
 DB 721 LPHSTPVSIDWILPPSEEFQVDAAPVPAPAPAGLPGVVLTPPPPVHKSIPSPS 780
 QY 781 RNRRLTYTPDGAKVYAGSLFESDCMLVNASNGHRPGGGLCHAFYQRPPEAFYPTFT 840
 DB 781 RNRRLTYTPDGAKVYAGSLFESDCMLVNASNGHRPGGGLCHAFYQRPPEAFYPTFT 840
 QY 841 MRGLAAYTLTPPIIHAVAPDVRVQGNPKRLAAYETOSRGTAYPLLSGSIYQVPV 900
 DB 841 MRGLAAYTLTPPIIHAVAPDVRVQGNPKRLAAYETOSRGTAYPLLSGSIYQVPV 900

DB 841 MRGLAAYTLTPPIIHAVAPDVRVQGNPKRLAAYETOSRGTAYPLLSGSIYQVPV 900
 QY 901 SLSPDAMERNHRPDELYLTPPEANMEANKPAPQVLTITTEDPRTANLLEIDAAEYVG 960
 DB 901 SLSPDAMERNHRPDELYLTPPEANMEANKPAPQVLTITTEDPRTANLLEIDAAEYVG 960
 QY 961 RACAGCTISPGIVHYOFTAGVPGSGKSRSLIOQGDVDVAVVPTRELRSWRRRGFAFTPH 1020
 DB 961 RACAGCTISPGIVHYOFTAGVPGSGKSRSLIOQGDVDVAVVPTRELRSWRRRGFAFTPH 1020
 QY 1021 TAARVTIGRRVVIDEAPSLPPLLHLMORASSVHLGDPNOIPADIDFBIAGLYPAIRPE 1080
 DB 1021 TAARVTIGRRVVIDEAPSLPPLLHLMORASSVHLGDPNOIPADIDFBIAGLYPAIRPE 1080
 QY 1081 LAPTSMMXVTHRCPADVCCELIRGAYPKIQTTSRVLSLFNNEPRIQOKLYXTOAKKAMP 1140
 DB 1081 LAPTSMMXVTHRCPADVCCELIRGAYPKIQTTSRVLSLFNNEPRIQOKLYXTOAKKAMP 1140
 QY 1141 GAITVHEAGATFETIITATDARGLIOSSRAHAVALTRHTEKCVILDPAGLLREVG1 1200
 DB 1141 GAITVHEAGATFETIITATDARGLIOSSRAHAVALTRHTEKCVILDPAGLLREVG1 1200
 QY 1201 SDVIYVNNFPLAGEVYKXHSVLRPGNPDONLTGLQAFPPSCQISAHQLAEBLGRHAP 1260
 DB 1201 SDVIYVNNFPLAGEVYKXHSVLRPGNPDONLTGLQAFPPSCQISAHQLAEBLGRHAP 1260
 QY 1261 VAAVLPCCPELEGGLYMPDELTVSDVLFELTDIVHCHMAAPSOGRKAVLSTLYGRYGR 1320
 DB 1261 VAAVLPCCPELEGGLYMPDELTVSDVLFELTDIVHCHMAAPSOGRKAVLSTLYGRYGR 1320
 QY 1321 RTKLYEAAHSDVRSARLFTPTIGPVQATTCELYELVEAAMVERGODGSAVLELDLCNRDV 1380
 DB 1321 RTKLYEAAHSDVRSARLFTPTIGPVQATTCELYELVEAAMVERGODGSAVLELDLCNRDV 1380
 QY 1381 SRITFEOKXCNKFTTGTETIAHGKVGQISAMSKTFCLPQWRAIKETLALLPPIEY 1440
 DB 1381 SRITFEOKXCNKFTTGTETIAHGKVGQISAMSKTFCLPQWRAIKETLALLPPIEY 1440
 QY 1441 GDAYEESVFAAASGAGSCAVFENDESEPDSTONNFSGLGECVMEBCGPMWLIRLYHL 1500
 DB 1441 GDAYEESVFAAASGAGSCAVFENDESEPDSTONNFSGLGECVMEBCGPMWLIRLYHL 1500
 QY 1501 VRSAMITIOAPKESLKGFWKHSGEPGTLIMNTYWNMAITAHCEFRDFRVAARKGDSDSV 1560
 DB 1501 VRSAMITIOAPKESLKGFWKHSGEPGTLIMNTYWNMAITAHCEFRDFRVAARKGDSDSV 1560
 QY 1561 LCSDYROSRAAALLIAGCGLKLVDYRPIGLYAGVVVAPGLTLPDVRFRAGRLSEKNMG 1620
 DB 1561 LCSDYROSRAAALLIAGCGLKLVDYRPIGLYAGVVVAPGLTLPDVRFRAGRLSEKNMG 1620
 QY 1621 PGBERAEQLRLAVCDPLRGTLTNVAQVCVDVVSRYVGPGLVHNLIGMLQTIADGKAHFT 1680
 DB 1621 PGBERAEQLRLAVCDPLRGTLTNVAQVCVDVVSRYVGPGLVHNLIGMLQTIADGKAHFT 1680
 QY 1681 ETIKPVLDLTNSIIORYE 1698
 DB 1681 ETIKPVLDLTNSIIORYE 1698

RESULT 2
 Q9YK10 PRELIMINARY; PRT; 1708 AA.
 AC Q9YK10;
 DT 01-MAY-1999 (Tremblrel. 10, Created)
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE NONSTRUCTURAL POLYPEPTIDE.
 OS Swine hepatitis E virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage.
 OX NCBI_TaxID=63421;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MENG;

RA MEDLINE=97420774; PubMed=9275216;
RA Meng X.J., Purcell R.H., Halbur P.G., Lehman J.R., Webb D.M.,
RA Tsareva T.S., Haynes J.S., Thacker B.J., Emerson S.U.;
RT "A novel virus in swine is closely related to the human hepatitis E
RT virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:9860-9865(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MENG.
RX MEDLINE=99030877; PubMed=9811705;
RA Meng X.J., Halbur P.G., Shapito M.S., Govindarajan S., Bruna J.D.,
RA Mushahar I.K., Purcell R.H., Emerson S.U.;
RT "Genetic and experimental evidence for cross-species infection by
RT swine hepatitis E virus.";
RL J. Virol. 72:9714-9721(1998).
DR EMBL: AF082843; AAC97208.1; -.
DR INTERPRO: IPR000606; -.
DR INTERPRO: IPR002588; -.
DR INTERPRO: IPR002589; -.
DR PFAM: PF01443; Viral helicase; 1.
DR PFAM: PF01660; Viral helicase; 1.
DR PFAM: PF01661; DUF27; 1.
KW Polyprotein.
FT CHAIN 55 237 METHYLTRANSFERASE.
FT CHAIN 975 1219 HELICASE.
FT CHAIN 1222 1708 RNA-DIRECTED RNA POLYMERASE.
SQ SEQUENCE 1708 AA; 187406 MW; 7444E52DCD616130 CRC64;

Query Match 97.8%; Score 8812.5; DB 14; Length 1708;
Best local similarity 97.8%; Pred. No. 0;
Matches 1661; Conservative 7; Mismatches 30; Indels 1; Gaps 1;

QY 1 PGTTAIEGAAALAAANSALANAVVPRFLSRVQTEILINLMOPRODVFREVLNHPHIO
DB 10 PGTTAIEGAAALAAANSALANAVVPRFLSRVQTEILINLMOPRODVFREVLNHPHIO
QY 61 VHNLEOYCRARAGCLVGAHPRISINPNVNLHRCFLRPVGRDQVRYSAFTRGPAAN
DB 70 AHNLEOYCRARAGCLVGAHPRISINPNVNLHRCFLRPVGRDQVRYSAFTRGPAAN
QY 121 CRRSALRGLPADRYTCFPGFSKCAFAETGVALYSLHDLMPADVAEAMARHGXTLYAA
DB 130 CRRSALRGLPADRYTCFPGFSKCAFAETGVALYSLHDLMPADVAEAMARHGXTLYAA
QY 181 LHLPEVLLPGTYHTSYLLIHDGDAVVTYEGDTSAQYNDVSLIRAMRTTKIVGDH
DB 190 LHLPEVLLPGTYHTSYLLIHDGDAVVTYEGDTSAQYNDVSLIRAMRTTKIVGDH
QY 241 PLVIEVRAIGCHFVLLTLAAPEPSPMPYVPRSTEVYVRSIFGCGSPSLPSCACSTK
DB 250 PLVIEVRAIGCHFVLLTLAAPEPSPMPYVPRSTEVYVRSIFGCGSPSLPSCACSTK
QY 301 STEFHAVPIHMDLMFGATLDDQAFCCSRIMTYLGRISYKTVGALVLANEGNASEDAL
DB 310 STEFHAVPIHMDLMFGATLDDQAFCCSRIMTYLGRISYKTVGALVLANEGNASEDAL
QY 361 TAVITAAVLTICHORYLRTQAISSKMRRLGVEIAOKFTIRLYSWLPEKSGRDVTPGROL
DB 370 TAVITAAVLTICHORYLRTQAISSKMRRLGVEIAOKFTIRLYSWLPEKSGRDVTPGROL
QY 421 FTAQCRMLISAGFHLDPRLVDESPYCRGRTFLKVAAGKFCCEFMWLGQECTFLPEAE
DB 430 FTAQCRMLISAGFHLDPRLVDESPYCRGRTFLKVAAGKFCCEFMWLGQECTFLPEAE
QY 481 GLVGDGHNEAVEGSEVDPAPBAHLDVSGTAVHGHOLEALYRALNVPODIAARASRLT
DB 490 GLVGDGHNEAVEGSEVDPAPBAHLDVSGTAVHGHOLEALYRALNVPODIAARASRLT
QY 541 ATVELVAPDRILECRTVLGNKTFRTTVVDGAHLLEANGPQDYVLSFPAASRQSGAGSHSLT
DB 550 ATVELVAPDRILECRTVLGNKTFRTTVVDGAHLLEANGPQDYVLSFPAASRQSGAGSHSLT

QY 601 YELTPAGLOVRISNGIDCTATPPGAPSAPEGEVAAPFASALYRYNFTQRHSLTGLM
DB 610 YELTPAGLOVRISNGIDCTATPPGAPSAPEGEVAAPFASALYRYNFTQRHSLTGLM
QY 661 LHPGGLGIFPPSPGHIMESANPCGEGTLTTRTSTSGSSDFSPPEAAMATPG 720
DB 670 LHPGGLGIFPPSPGHIMESANPCGEGTLTTRTSTSGSSDFSPPEAAMATPG 729
QY 721 LHPSTPEVSDIWLPPPESEFOYDAAPVPAPDPAGLPGCVUT-PPPPVYKPSITPP 779
DB 730 LHPSTPEVSDIWLPPPESEFOYDAAPVPAPDPAGLPGCVUT-PPPPVYKPSITPP 789
QY 780 SRRRLTYTPDGAKYVAGSLFESDDMWLVNANPGHRRPGGGLCHAFYORFPAPFPTER 839
DB 790 SRRRLTYTPDGAKYVAGSLFESDDMWLVNANPGHRRPGGGLCHAFYORFPAPFPTER 849
QY 840 IMREGLAAYTLTPRIIHAAPDYRYEONKRLAEAYRETCSSRRGTAAAYLLSGIYQV 899
DB 850 IMREGLAAYTLTPRIIHAAPDYRYEONKRLAEAYRETCSSRRGTAAAYLLSGIYQV 909
QY 900 VLSLSPDAMERNHRPGDELTYTEPAANFEANKPAQVLTITEDTARTANALEIDAATEV 959
DB 910 VLSLSPDAMERNHRPGDELTYTEPAANFEANKPAQVLTITEDTARTANALEIDAATEV 969
QY 960 GRACAGCTSPGIVHYOFTAGVPGSGKRSIOGDVYVVPTRRLNSRRRGFAFTP 1019
DB 970 GRACAGCTSPGIVHYOFTAGVPGSGKRSIOGDVYVVPTRRLNSRRRGFAFTP 1029
QY 1020 HTAARVYIGRRVYIDEAPSLPHLLLLHMQRASSVHLGDPNOIPADIEFHAGLVPAIRP 1079
DB 1030 HTAARVYIGRRVYIDEAPSLPHLLLLHMQRASSVHLGDPNOIPADIEFHAGLVPAIRP 1089
QY 1080 ELAPTSMWXYTHRCPADVCELIRGAYPKIOTTSRVLSLFWNEBPAIGOKLVYTOAKAAN 1139
DB 1090 ELAPTSMWXYTHRCPADVCELIRGAYPKIOTTSRVLSLFWNEBPAIGOKLVYTOAKAAN 1149
QY 1140 PGATVHEOGAFTETTTIATADARGLIOSSRAHAVALTRTEKCVIADAGLLREVG 1199
DB 1150 PGATVHEOGAFTETTTIATADARGLIOSSRAHAVALTRTEKCVIADAGLLREVG 1209
QY 1200 ISDVIANNEFLAGGEVXHRPSYIPRGNPDONIGTLOAFPPSCQISAYHOLAEELGRPA 1259
DB 1210 ISDVIANNEFLAGGEVXHRPSYIPRGNPDONIGTLOAFPPSCQISAYHOLAEELGRPA 1269
QY 1260 PVAAYLPCELEBQGLLYMPQELTVSDSVLVELTDLVHCMAAPSQKAVLSTLVGRYG 1319
DB 1270 PVAAYLPCELEBQGLLYMPQELTVSDSVLVELTDLVHCMAAPSQKAVLSTLVGRYG 1329
QY 1320 RRTKLYEAASDVRESLARIPTIGVQATTCELYELVEMVNEKGGDGSVLELDCNRD 1379
DB 1330 RRTKLYEAASDVRESLARIPTIGVQATTCELYELVEMVNEKGGDGSVLELDCNRD 1389
QY 1380 VSRITFEOKXCNKFTTGETTAHKGVGGISAMSKTFCCALGPMFRAIEKILALPNIF 1439
DB 1390 VSRITFEOKXCNKFTTGETTAHKGVGGISAMSKTFCCALGPMFRAIEKILALPNIF 1449
QY 1440 YGDAYEESVFAAASGAGSCMVEENDESEFSTONNESLGLCECYVMECGMPQLIRLYH 1499
DB 1450 YGDAYEESVFAAASGAGSCMVEENDESEFSTONNESLGLCECYVMECGMPQLIRLYH 1509
QY 1500 LVRSANILQAPKESLKGFMKKHSGEPGTLMTNVMNNAIIAHQYEPFDFVAAFKGGDSV 1559
DB 1510 LVRSANILQAPKESLKGFMKKHSGEPGTLMTNVMNNAIIAHQYEPFDFVAAFKGGDSV 1569
QY 1560 VLSGDSYRQSNAAALJAGGLKLVYRPIGLAGVYVAPAGLGLTPVYVFAARLSKKNN 1619
DB 1570 VLSGDSYRQSNAAALJAGGLKLVYRPIGLAGVYVAPAGLGLTPVYVFAARLSKKNN 1629
QY 1620 GPGPERAEOQLAVCDLRLGTLNVAQVYDVYVSVGSLVHNLIGMLQTTADGKAHF 1679
DB 1630 GPGPERAEOQLAVCDLRLGTLNVAQVYDVYVSVGSLVHNLIGMLQTTADGKAHF 1689
QY 1680 TETIKPVLDLTNIIORVE 1698

DB	1690	TETIKPVLDTINSTIQRVE	1708
RESULT	3		
Q9YLRL	Q9YLRL	PRELIMINARY;	PRT; 1708 AA.
AC	Q9YLRL		
DT	01-MAY-1999	(TREMBLrel. 10, Created)	
DT	01-MAY-1999	(TREMBLrel. 10, Last sequence update)	
DE	01-JUN-2000	(TREMBLrel. 14, Last annotation update)	
DE	POLYPROTEIN.		
GN	ORF1.		
OS	Hepatitis E virus.		
OC	Viruses: ssRNA positive-strand viruses, no DNA stage.		
OX	NCBI_TaxID=12461;		
RN	[1]		
RC	SEQUENCE FROM N.A.		
RC	STRAIN-HEV-US2;		
RX	MEDLINE=98178637; PubMed=9519822;		
RA	Schlauder G.G., Dawson G.J., Erker J.C., Kwo P.Y., Koluge M.F., Smalley D.L., Rosenblatt J.E., Desai S.M., Mushahwar I.K., "The sequence and phylogenetic analysis of a novel hepatitis E virus isolated from a patient with acute hepatitis reported in the United States.";		
RT	J. Gen. Virol. 79:0-0(0).		
RL	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-HEV-US2;		
RX	MEDLINE=99190429; PubMed=10092008;		
RA	Erker J.C., Desai S.M., Schlauder G.G., Dawson G.J., Mushahwar I.K., "A hepatitis E virus variant from the United States: molecular characterization and transmission in cynomolgus macaques.";		
RT	J. Gen. Virol. 80:681-690(1999).		
RL	EMBL: AF060669; AAD15815.1; -		
DR	INTERPRO: IPR000606; -		
DR	INTERPRO: IPR002588; -		
DR	INTERPRO: IPR002589; -		
DR	PFAM: PF01443; Viral_helicase1; I.		
DR	PFAM: PF01660; Methyltransf; 1.		
DR	PFAM: PF01661; DUF27; 1.		
KW	Polyprotein.		
SEQ	SEQUENCE 1708 AA; 187260 MW; BE0DF5BEEDAB547 CRC64;		
Query Match	97.4%; Score 8780.5; DB 14; Length 1708;		
Best Local Similarity	97.4%; Pred. No. 0;		
Matches 1655; Conservative	8; Mismatches 35; Indels 1; Gaps 1;		
QY	1	PGTTAIEOALAAANSALANAVVVRPELSKVOTELINLMOPROLVREPVLMNHPIOR	60
DB	10	PGTTAIEOALAAANSALANAVVVRPELSKVOTELINLMOPROLVREPVLMNHPIOR	69
QY	61	VHNELEQYCRARARCLVEGAHPRSINDNPNVLRHCFLRPGVQRYRYSAPTRGPAN	120
DB	70	VHNELEQYCRARARCLVEGAHPRSINDNPNVLRHCFLRPGVQRYRYSAPTRGPAN	129
QY	121	CRSALRGLPADRYCEFDGFSRCAFAAETGALYSLHLMVPADVAEAMARAGXRLTAA	180
DB	130	CRSALRGLPADRYCEFDGFSRCAFAAETGALYSLHLMVPADVAEAMARAGXRLTAA	189
QY	181	LHLPEVLLPGTYTTSYLLIHGDRAVYVYEGDTSAGYNHDSVILRAWITRTKIAGDH	240
DB	190	LHLPEVLLPGTYTTSYLLIHGDRAVYVYEGDTSAGYNHDSVILRAWITRTKIAGDH	249
QY	241	PLVERVAIGCHEVLLTAAPEBPMYVVRPRSTEVYVRSIEBPGGSPSLFPACSTK	300
DB	250	PLVERVAIGCHEVLLTAAPEBPMYVVRPRSTEVYVRSIEBPGGSPSLFPACSTK	309
QY	301	STFNAVPHIMDRLMFGATLDDQAFCCSRMLTMYRGISYKTVGALVANEGMNASEDAL	360
DB	310	STFNAVPHIMDRLMFGATLDDQAFCCSRMLTMYRGISYKTVGALVANEGMNASEDAL	369

Qy	361	TAXITAAVLTI	CHORYLRTQIAISKGMRLGVEHAQEFITRLYSWLEPESGRVY	IGRDLQ	420
Dd	370	TAVITAAVLTI	CHORYLRTQIAISKGMRLGVEHAQEFITRLYSWLEPESGRVY	IGRDLQ	429
Qy	421	FYAQCRWLSAGFHL	DBRVLVFDESVCRCRTRFLKLVAAKFCCEFMRLGDECTFLPAE		480
Dd	430	FYAQCRWLSAGFHL	DBRVLVFDESVCRCRTRFLKLVAAKFCCEFMRLGDECTFLPAE		489
Qy	481	GLVGDHGDNA	YATGSEYDAEPANHLDSQTYAVBHQLEALYALANPQDIAARASLT		540
Dd	490	GLVGDHGDNA	YATGSEYDAEPANHLDSQTYAVBHQLEALYALANPQDIAARASLT		549
Qy	541	ATVELVASPDL	ECRTVLGNKTFPTTYVDCANLEANGPROYVLSFSDASROSAGASHLT		600
Dd	550	ATVELVASPDL	ECRTVLGNKTRFTTYVDCANLEANGPEYVLSFSDASROSAGASHLT		609
Qy	601	YELTPAGLOVR	ISSNGDLCTATPPGGAPSAAGVEAAACSAALYNNFTQHNHSLTGJM		660
Dd	610	YELTPAGLOVR	ISSNGDLCTATPPGGAPSAAGVEAAACSAALYNNFTQHNHSLTGJM		669
Qy	661	LHPEGLIGI	FPFSPGHIWESANPFCGEGTLYTRWTSJGFSDDSPPEAAAPAMAAATPG		720
Dd	670	LHPEGLIGI	FPFSPGHIWESANPFCGEGTLYTRWTSJGFSDDSPPEAAAPASAAAPG		729
Qy	721	LPHSTPPVSDI	WVLPSPSEEOQDAAPVPARPACLTBPVYLT-PPRPYHNKSP		779
Dd	730	LPHSTPPVSDI	WVLPSPSESHVDASVPSPACLTSPYVLTpppppppRRKPAISPP		789
Qy	780	SHNRRLTY	YPDGAKYVAGSLJFESDCDWLVANSPNHRGGGLCHAFYOREPEAFYPRFE		839
Dd	790	PTRRLTY	YPDGAKYVAGSLXESDCDWLVANSPNHRGGGLCHAFYOREPEAFYSTEF		849
Qy	840	IMREGLA	YTLTPRPITHAAPDYRVQONKRLAAAYRETCSRGTAAYPLLGSIYVP		899
Dd	850	IMREGLA	YTLTPRPITHAAPDYRVQONKRLAAAYRETCSRGTAAYPPLGSGIYVP		909
Qy	900	VLSFSDAERNHR	RGDELYLTERPAMNFEANKPAQVLTITDTRTANLLEIDAAREV		959
Dd	910	VLSFSDAERNHR	RGDELYLTERPAMNFEANKPAQVLTITDTRTANLLEIDAAREV		969
Qy	960	GRACAGCTI	SPGIYHOFYTAGVPDSGKSRSIQGGDVVVVYPTRELNSWRRRGFAETP		1019
Dd	970	GRACAGCTI	SPGIYHOFYTAGVPDSGKSRSIQGGDVVVVYPTRELNSWRRRGFAETP		1029
Qy	1020	HTAAVYTI	GRNVYIDEAPSLPRHLLLNHRASSVNLGDPQITADIFENAGLYPAIRP		1079
Dd	1030	HTAAVYTI	GRNVYIDEAPSLPRHLLLNHRASSVNLGDPQITADIFENAGLYPAIRP		1089
Qy	1080	ELAPRSMXV	YHNRPAVCELRGAYPKIOTTSRYVLSLFEMNEPRLGOKLVYXTOAKKAN		1139
Dd	1090	ELAPRSMXV	YHNRPAVCELRGAYPKIOTTSRYVLSLFEMNEPRLGOKLVYXTOAKKAN		1149
Qy	1140	PGAITVHEA	QAGATFETTITATADARGLIOSSRAHAIVALJTRHTEKCYILDAFGJLEVG		1199
Dd	1150	PGAITVHEA	QAGATFETTITATADARGLIOSSRAHAIVALJTRHTEKCYILDAFGJLEVG		1209
Qy	1200	ISDVIYNNF	ELAGEVGHNRPSVPRGNPDQNLGTLOARPSCQISAYHQLAEELGHRPA		1259
Dd	1210	ISDVIYNNF	ELAGEVGHNRPSVPRGNPDQNLGTLOARPSCQISAYHQLAEELGHRPA		1269
Qy	1260	PYAANLP	PCPELEBGLYMQDELTSVSYVLEFLDTIYICRMAASOKKAVLSTLYGKG		1319
Dd	1270	PYAANLP	PCPELEBGLYMQDELTSVSYVLEFLDTIYICRMAASOKKAVLSTLYGKG		1329
Qy	1320	RTTKLYEASH	DVEESLARFPTIGPVQATCGLYELVAMYEKQODSAYLEIDLNRD		1379
Dd	1330	RTTKLYEASH	DVEESLARFPTIGPVQATCGLYELVAMYEKQODSAYLEIDLNRD		1389
Qy	1380	VSRTIFPEAK	XKNKFTTJETITAHKVOQGISANSKTEFCALFGWPAIRAIKEITALLPRNIF		1439
Dd	1390	VSRTIFPEAK	XKNKFTTJETITAHKVOQGISANSKTEFCALFGWPAIRAIKEITALLPRNIF		1449
Qy	1440	YGDAEBSV	FPAAYVAGSGCMVEENDFSEFDTQNNFSLGLECYVMEBCGMPQWLLIRLYH		1499

```

Db 1450 YGDAYEESVFAAASGASGCMVFENDESEFDSTQNNFSLGLECVMEECGMQMLRLRH 1509
QY 1500 LVRSANLIDAPKESLKGFKKHSGEPTLLMNTVMNMAIIACYERDRFVAAFKDDSV 1559
Db 1510 LVRSANLIDAPKESLKGFKKHSGEPTLLMNTVMNMAIIACYERDRFVAAFKDDSV 1569
QY 1560 VICSVDROSMAALTAGGGLKTKVYRPIGLYAGVVAAGLTPDDVYRFGARLSEKNW 1619
Db 1570 VICSVDROSMAALTAGGGLKTKVYRPIGLYAGVVAAGLTPDDVYRFGARLSEKNW 1629
QY 1620 GGPBERAEDRLAVCDLFLGLTNVAQVYVSRVYSGPLVHNILGMLQTTADGKAH 1679
Db 1630 GGPBERAEDRLAVCDLFLGLTNVAQVYVSRVYSGPLVHNILGMLQTTADGKAH 1689
QY 1680 TETIKPVDLINSIIORVE 1698
Db 1690 TENIKPVDLINSIIORVE 1708

RESULT 4
Q9IVZ9 PRELIMINARY; PRT; 1707 AA.
AC Q9IVZ9;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE POLYPROTEIN.
OS Hepatitis E virus:
OC Viruses: ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-T1;
RA Harrison T.J.;
RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-T1;
RA Wang Y., Zhang H., Ling R., Li H., Harrison T.J.;
RT "The complete sequence of hepatitis E virus genotype 4 reveals an
RT alternative strategy for translation of open reading frames 2 and 3."
RL J. Gen. Virol. 81:1675-1686(2000).
DR EMBL: AJ272108; CAB83209.1; -.
SQ SEQUENCE 1707 AA; 187296 MW; 2A80F303AB98C4EC CRC64;

Query Match 85.8%; Score 7730; DB 14; Length 1707;
Best Local Similarity 83.9%; Pred. No. 0;
Matches 1444; Conservative 91; Mismatches 140; Indels 46; Gaps 5;
QY 1 PGTTAIEAOALAAANSALANAVVAPFLSRVOTELLINLMOPROLVFREPVLNHPRIOR 60
Db 10 PGVTAIEAOALAAANSALANAVVAPFLSRVOTELLINLMOPROLVFREPVLNHPRIOR 69
QY 61 VHNLEEOYCRARAGRCLEVGAPRSINPNVLRHCFLRPGVRDVOHWASPTRGPAAN 120
Db 70 VHNLEEOYCRARAGRCLEVGAPRSINPNVLRHCFLRPGVRDVOHWASPTRGPAAN 129
QY 121 CRRSALRGLPADRTYCFDGFSCAFEAETGVALYSLHDLMPADVAEAMARHGXTPLUYA 180
Db 130 CRRSALRGLPADRTYCFDGFSCAFEAETGVALYSLHDLMPADVAEAMARHGXTPLUYA 189
QY 181 LHLPEPVLPLPGTYHTTSTLLYHGDRAVYTYEGDTSAGYNHDSILRAMIRTTKTVGDH 240
Db 190 LHLPEPVLPLPGTYHTTSTLLYHGDRAVYTYEGDTSAGYNHDSILRAMIRTTKTVGDH 249
QY 241 PLVIERVARIGCHFVLLTAAPESPMPVVPYPRSTREYVRSIFEGGSPSLPSPASCNRK 300
Db 250 PLVIERVARIGCHFVLLTAAPESPMPVVPYPRSTREYVRSIFEGGSPSLPSPASCNRK 309
QY 301 STEHAHVHIMDRMLFGATLDDQAFCCSRLMTYLRGISYKTVGALVANEGNASEDAL 360

```

```

Db 310 STEHAHVHIMDRMLFGATLDDQAFCCSRLMTYLRGISYKTVGALVANEGNASEDAL 369
QY 361 TAXITAAVYLICQRIYLRTOAISKMRILGVEHAQKFITRILYSWLEFKSRDYIPGROLO 420
Db 370 TAXITAAVYLICQRIYLRTOAISKMRILGVEHAQKFITRILYSWLEFKSRDYIPGROLO 429
QY 421 FYAQCRRLWSAGHLLDPRLVYVDESVCRCRTEFLKKYVAGFECGFMRLGQECFELPAE 480
Db 430 FYAQCRRLWSAGHLLDPRLVYVDESVCRCRTEFLKKYVAGFECGFMRLGQECFELPAE 489
QY 481 GLVGDGHNDNEAEGSEVDPAPBAHLDVSGTAVVHGHOLEALYRALNVDPDIAARASRLT 540
Db 490 GRVGEQYIDENAEFGSDIDPAEEATYSIAGSYLVTSQOLPLQALGIPSDLAARASRLT 549
QY 541 ATVELVASPDRLCEKTVLGKTKTEFTTVVDGAHLLEANGPEQYVLSFDPASRQSMGASHSLT 600
Db 550 ATVELVASPDRLCEKTVLGKTKTEFTTVVDGAHLLEANGPEQYVLSFDPASRQSMGASHSLT 609
QY 601 YELTPAGLOVYRISNGLDCTATFPFGAPSAAGCEVAAPCSALYXRNRFQORSLTGLM 660
Db 610 YELTPAGLOVYRISNGLDCTATFPFGAPSAAGCEVAAPCSALYXRNRFQORSLTGLM 669
QY 661 LHPEGLGIFPPSPGHIWESANPCGEGTLYTRTSTSGFSSDFSPPEAARAPAMATPG 720
Db 670 YHPEGLVGLFPFSPGHIWESANPCGEGTLYTRTSTSGFSSDFSPPEAARAPAMATPG 724
QY 721 LPHSTPPVSDIWLPPSE--EFQYDAP-----VPPADPAGLP----- 758
Db 725 -----PPAEVNTPEVLDALPSEIMERPAPASBPAPSPSDVNSFPTS 770
QY 759 -GPVLTLPFPPPVYKPSLPPPSRNRLTYTDGAKVYVAGSLFESCDMLVYVANSNGRH 817
Db 771 SGAPLAPAPALPVLTHLSCP---RRLLTYTDGSGVYVAGSLFESCDMLVYVANSNGRH 826
QY 818 PGGLCHAFYORPEAFYEFITMRGLAAYTLPRIHAVAVDVVEONPRLAAYR 877
Db 827 PGGLCHAFYORPEAFYEFITMRGLAAYTLPRIHAVAVDVVEONPRLAAYR 886
QY 878 ETGSRGTAAYPRLIGSIVOVVPSLSFDAMERNHRPDELTYLEPAMNEANKPAOPVL 937
Db 887 ETGSRGTAAYPRLIGSIVOVVPSLSFDAMERNHRPDELTYLEPAMNEANKPAOPVL 946
QY 938 TTEPDARTANLALIEDAENVRACAGCTISGCIHYOCTAACPVGSKRSIQOQDVY 997
Db 947 TTEPDARTANLALIEDAENVRACAGCTISGCIHYOCTAACPVGSKRSIQOQDVY 1006
QY 998 VVVPTRRLNSWRMRGFAFTPHTAARTIGRNVVIDEASLPPHLLLMORASSVHL 1057
Db 1007 IVPTRRLNSWRMRGFAFTPHTAARTIGRNVVIDEASLPPHLLLMORASSVHL 1066
QY 1058 GPNQIPADIDFENAGLVPATRPDLAPTSWXYVTHRCPADVCELRGAYPKIOTTSRYLRS 1117
Db 1067 GPNQIPADIDFENAGLVPATRPDLAPTSWXYVTHRCPADVCELRGAYPKIOTTSRYLRS 1126
QY 1118 LFMNEPAIGOKLYXYOAAKANKANGAITVHQAQATFETTTIATAARGLIQSSRAAYI 1177
Db 1127 LFMNEPAIGOKLYXYOAAKANKANGAITVHQAQATFETTTIATAARGLIQSSRAAYI 1186
QY 1178 ALTRHTEKCIIDAPGLREVGISDVIVNNFPLAGEVGXHRPSVYPRGNPDNLGTLQA 1237
Db 1187 ALTRHTEKCIIDAPGLREVGISDVIVNNFPLAGEVGXHRPSVYPRGNPDNLGTLQA 1246
QY 1238 FPPSCQISAVHQAELGHRPAPVAVLPPCPLEBGLYMPDELYSDSVYFELTDIV 1297
Db 1247 FPPSCQISAVHQAELGHRPAPVAVLPPCPLEBGLYMPDELYSDSVYFELTDIV 1306
QY 1298 HCMAPASORKAVALSTLVGYGRTKLYEASHDVDESRLAPFTPTGPOVATCELYEY 1357
Db 1307 HCMAPASORKAVALSTLVGYGRTKLYEASHDVDESRLAPFTPTGPOVATCELYEY 1366
QY 1358 EAMVEKGQDGSAYVELDLNDRYSRTTFQKCNKFTTGTGTTIAGHGVQGISAMSKTFCA 1417
Db 1367 EAMVEKGQDGSAYVELDLNDRYSRTTFQKCNKFTTGTGTTIAGHGVQGISAMSKTFCA 1426

```

QY 1418 LFGFWFAIEKEILLALPPNIFYGDAYEESYFAAASGAGSCWVENDFSEFDSTONNES 1477
1427 LFGFWFAIEKEILLALALPNVIFYGDAYEDVLAFAVAGAPCKVFNDESEFDSTONNES 1486
QY 1478 LGLCEVYMEGCGQWILRLYLHLYRSAMILLQAPKESLKGFKKHSRGTLLMNTVMNMA 1537
1487 LGLCEIIMEGCGQWILRLYLHLYRSAMILLQAPKESLKGFKKHSRGTLLMNTVMNMA 1546
QY 1538 IIAHCEYFDFRVAFAFGDSDSVLCSYRQSRNAALAGCGLKLYDPRIGLYAGVYV 1597
1547 VIAHCYFDFRDLKVAFAFGDSDSVLCSYRQSRNAALAGCGLKLYDPRIGLYAGVYV 1606
QY 1598 APGLGLTLDVYVFAFGRLSEKKNMGPPERAQDLAVCDLGLTNVAQCVVVSRYGV 1637
1607 APGLGLTLDVYVFAFGRLSEKKNMGPPERAQDLAVCDLGLTNVAQCVVVSRYGV 1666
QY 1658 SPGLVHNLIGLQTIADGKAHFTETIKPVLDLTIISIIQRE 1698
1667 SPGLVHNLIGLQTIADGKAHFTETIKPVLDLTIISIIQRE 1707

RESULT 5
069410
ID 069410 PRELIMINARY; PRT: 1693 AA.
AC 069410;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE METHYL TRANSFERASE.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
NCBI_TaxID=12461;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=HEV037;
RA Donati M.C., Fagan E.A., Harrison T.J.;
RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL: X88292; CAAG6936.1; -
DR INTERPRO: IPR000606; -
DR INTERPRO: IPR002588; -
DR INTERPRO: IPR002589; -
DR PRAM: PF01443; Viral_helicase1; 1.
DR PFAM: PF01660; Vmethyltransf; 1.
DR PFAM: PF01661; DUF27; 1.
KW Transferrase.
SQ SEQUENCE 1693 AA; 185190 MW; C347C2436DED9B23 CRC64;

Query Match 82.9%; Score 7468; DB 14; Length 1693;
Best Local Similarity 82.3%; Pred. No. 0;
Matches 1402; Conservative 109; Mismatches 167; Indels 26; Gaps 7;

QY 1 PGTITAEQAALAAANALANAVVVRFLSRVQTEIILINLMOPQOLVFRREVLMNHQIQR 60
10 PGTITAEQAALAAANALANAVVVRFLSRVQTEIILINLMOPQOLVFRREVLMNHQIQR 69
QY 61 VINHELEQYCRABAGRCLEVGARHSINDNPNVLRCELRPVGDVORWYSAPTRGPAAN 120
70 VINHELEQYCRABAGRCLEVGARHSINDNPNVLRCELRPVGDVORWYSAPTRGPAAN 129
QY 121 CRSSALRGLEPPADRTYCEDFSRCAFAETGVALYSLHDLMPADVADAMARHGXTRLYAA 180
130 CRSSALRGLEPPADRTYCEDFSRCAFAETGVALYSLHDLMPADVADAMARHGXTRLYAA 189
QY 181 LHPPEVLLPRTGYHTTSYLLIHGDRVAVTYEGDTSAGTNHDYSILIRAITRTKIYGDH 240
190 LHPPEVLLPRTGYHTTSYLLIHGDRVAVTYEGDTSAGTNHDYSILIRAITRTKIYGDH 249
QY 241 PLVIERAIGCHFVLLITAPAPSPMPVYVPRSTEVYRSIFGPGTSLPFTSCSTK 300
250 PLVIERAIGCHFVLLITAPAPSPMPVYVPRSTEVYRSIFGPGTSLPFTSCSTK 309

QY 301 STEFAVPHVIMDRMLFGLATLDQAFCSSRLMTYLRGISTYKVTVGALVANEGMNASDAL 360
310 STEFAVPHVIMDRMLFGLATLDQAFCSSRLMTYLRGISTYKVTVGALVANEGMNASDAL 369
QY 361 TAXITAAVLTICHOHYRLTQATISGMRRLVEHOKETITRLYSWLFKSGSDYIRGQLO 420
370 TAXITAAVLTICHOHYRLTQATISGMRRLVEHOKETITRLYSWLFKSGSDYIRGQLO 429
QY 421 FYAOCRRMLISAGFHLDRVVLFEDESVPQRCFTFLKKAAGKFCGEMRLGECSTFLPAE 480
430 FYAOCRRMLISAGFHLDRVVLFEDESVPQRCFTFLKKAAGKFCGEMRLGECSTFLPAE 489
QY 481 GLVGDHNDNEAYEGSEVDAPRAHLVDYSGTYAVHQLLEALYALANVPDIIARASRLT 540
490 GLVGDHNDNEAYEGSEVDAPRAHLVDYSGTYAVHQLLEALYALANVPDIIARASRLT 549
QY 541 ATVELVASPDLRECRVYLGKNTFRTVVDGHLBANPQEVYLFSDASRQMGASLSLT 600
550 ATVELVASPDLRECRVYLGKNTFRTVVDGHLBANPQEVYLFSDASRQMGASLSLT 609
QY 601 YELTPAGLQVRISSNGLDCTATPPPGAPSAAPGEVAFAFCALRYNRFTQRIHSLTGLW 660
610 YELTPAGLQVRISSNGLDCTATPPPGAPSAAPGEVAFAFCALRYNRFTQRIHSLTGLW 669
QY 661 LHPGLGLITPPSPGRIWESANPFCGEGTYLTYTWS-TSGFSSDSP-----PEAAPA 714
670 LHPGLGLITPPSPGRIWESANPFCGEGTYLTYTWS-TSGFSSDSP-----PEAAPA 729
QY 715 MAATPGLPHSNPPVSDLWVLPPESEPOVDAAPVPAAPDAGLGPVLPPEPPHPK 774
730 MAATPGLPHSNPPVSDLWVLPPESEPOVDAAPVPAAPDAGLGPVLPPEPPHPK 773
QY 775 SIIPPSRNRLLTYTPDGAAYAGSLFESDCDWLVNANPGRHGGGLCHAFYORPEAF 834
774 SIIPPSRNRLLTYTPDGAAYAGSLFESDCDWLVNANPGRHGGGLCHAFYORPEAF 829
QY 835 YPTEFIMREGIATTLTPRPIIAHAVADRYEONPKLEAAYRTGCRGTAATPLILGSG 894
830 YPTEFIMREGIATTLTPRPIIAHAVADRYEONPKLEAAYRTGCRGTAATPLILGSG 889
QY 895 IYQVPLSFDAMERNRPPGDELTYLPEANWFEANPAPVLTITDTRTAVLALCID 954
890 IYQVPLSFDAMERNRPPGDELTYLPEANWFEANPAPVLTITDTRTAVLALCID 949
QY 955 AATEVGRACAGCTISPGIVHYOFTAGVPGSGKSRISIOGDVVDVVTRELNRMSRRRGF 1014
950 AATEVGRACAGCTISPGIVHYOFTAGVPGSGKSRISIOGDVVDVVTRELNRMSRRRGF 1009
QY 1015 AATPHTAARVTTIGRRVVIDBPSLPHLLILHMQRASSVHLGDDPNQIPALIDEHAGLV 1074
1010 AATPHTAARVTTIGRRVVIDBPSLPHLLILHMQRASSVHLGDDPNQIPALIDEHAGLV 1069
QY 1075 PAIRPELAPTSMXVYHRCADVCCELIRGAVPKIOTSRVLSLSENEPRAIGOKLVYQA 1134
1070 PAIRPELAPTSMXVYHRCADVCCELIRGAVPKIOTSRVLSLSENEPRAIGOKLVYQA 1129
QY 1135 AKANPAITVHEAGATFETITATADARGLIOSSRAHAIALYALHTTEKCVILDAAGL 1194
1130 AKANPAITVHEAGATFETITATADARGLIOSSRAHAIALYALHTTEKCVILDAAGL 1189
QY 1195 LREYGISDVIVNFFLAGGEVGHKRSVIFRGNDONLGLTLOAFPPSCQISAYHQLAEEL 1254
1190 LREYGISDVIVNFFLAGGEVGHKRSVIFRGNDONLGLTLOAFPPSCQISAYHQLAEEL 1249
QY 1255 GHRPAPVAAVLPQCELEGLLVMPOSLYSDSVLVELLNDIYHCRMAASQKRAVISTL 1314
1250 GHRPAPVAAVLPQCELEGLLVMPOSLYSDSVLVELLNDIYHCRMAASQKRAVISTL 1309
QY 1315 VGRYGRRTKLYEAAHSDVRSLSARFIPTIGPVQATTCELEIYEAAMEKQODSAGVLELD 1374
1310 VGRYGRRTKLYEAAHSDVRSLSARFIPTIGPVQATTCELEIYEAAMEKQODSAGVLELD 1369
QY 1375 LCNBDVSRITFEOKXCKNFTTGETTIAGHKVGGCISAMSKTFICALGFWPRAIEKEILLAL 1434

```

Db 1370 LCNKRDVSRITFTFOCKCNKTTGTTGTAHGKVGOCISAMSTFCALFEPWRAITAKALLALL 1429
Qy 1435 PNIFYGADYEESVFAAAYSGAGSCWYFENDESEFDSTONNFSLSGECVYMECGNPOWL 1494
Db 1430 PQGVFEFGADFTVFAAATAAASAMVAFENDESEFDSTONNFSLSGECALMECGNPOWL 1489
Qy 1495 IRLYHLVRSAMTLOAKRESLSKGFWMKHSSEPGTLLMNTYMNNAITAHCEYFRFRRAAKR 1554
Db 1490 IRLYHLIRSAMTLOAKRESLSKGFWMKHSSEPGTLLMNTYMNNAITAHCEYFRFRRAAKR 1549
Qy 1555 GDDSVYLCSDYRSMNAALILAGCGKLKYDRPIGLYGVVYAPGTLTLDVYVFRAGRL 1614
Db 1550 GDDSVYLCSEYRSPCAATVLIAGCGKLKYDRPIGLYGVVYAPGTLTLDVYVFRAGRL 1609
Qy 1615 SEKNMGPGGERAEQRLAVCDLRLGTLNVAQVCVDVYSRVYSGPLVHNLIGMLQTIAD 1674
Db 1610 TEKMGPGGERAEQRLAVSDFLRLKLTNVAQVCVDVYSRVYSGPLVHNLIGMLQTIAD 1669
Qy 1675 GRAHPTETIKPVLDLTNLTIORE 1698
Db 1670 GRAHPTESVAPVLDLTNLTICRAVE 1693

RESULT 6
Q9WLL5 PRELIMINARY: PRT: 1693 AA.
ID Q9WLL5;
AC Q9WLL5;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDE.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TK15/92;
RA MEDLINE=99049628; PubMed=9833882;
RX Gouvea V., Snellings N., Popek M.J., Longer C.F., Innis B.L.;
RT "Hepatitis E virus: complete genome sequence and phylogenetic analysis
of a Nepal isolate."
RT of a Nepal isolate."
RL Virus Res. 57:21-26(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TK15/92;
RA Gouvea V.;
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF051830; AAC97186.1; -
DR INTERPRO: IPR000606; -
DR INTERPRO: IPR002588; -
DR INTERPRO: IPR002589; -
DR PFAM: PF01443; Viral_helicase1.1.
DR PFAM: PF01660; Vmchyltransf.1.
DR PFAM: PF01661; DUF27.1.
KW Polyprotein.
SQ SEQUENCE 1693 AA: 185349 MW: 4895ACDIDAEZFBBD CRC64:

```

Query Match 82.7%; Score 7457; DB 14; Length 1693;
 Best Local Similarity 81.7%; Pred. No. 0;
 Matches 1397; Conservative 110; Mismatches 166; Indels 36; Gaps 6;

```

Db 130 CRSSALRGLEPADRTYCDFDSCGSPPEPENGIALSLHDMSPDVAEAMRHGMRTLXAA 189
Qy 181 LHLPEVLLPPGTYHTTYSYLLHDDGRAVVTYEGDTSAGYNHDVSLIRANIRTKKIVGDH 240
Db 190 LHLPEVLLPPGTYHTTYSYLLHDDGRAVVTYEGDTSAGYNHDVSLIRANIRTKKIVGDH 249
Qy 241 PLVIERVRAIGCHFVLLTLTAAPESPMPYVPYRSTREYVYRSIFGGSGSLTSPSACSTK 300
Db 250 PLVIERVRAIGCHFVLLTLTAAPESPMPYVPYRSTREYVYRSIFGGSGSLTSPSACSTK 309
Qy 301 STFAHVAHVHIDMLMFGATLDDQAFCCSRLMTYLGISKVVYAGLVANEGNNASEDAL 360
Db 310 STFAHVAHVHIDMLMFGATLDDQAFCCSRLMTYLGISKVVYAGLVANEGNNASEDAL 369
Qy 361 TAXITAAVLTTCQRYLRTQAIKSKMRRLGVEHAQKFTRLYSWLFKSGRDYIPGRQLO 420
Db 370 TAXITAAVLTTCQRYLRTQAIKSKMRRLGVEHAQKFTRLYSWLFKSGRDYIPGRQLO 429
Qy 421 FYAQCRRLWSAGFHLDPRVLYPDESVPCCRTRFLKKVAGKFCGEMRLGECFCLEPAE 480
Db 430 FYAQCRRLWSAGFHLDPRVLYPDESAPCHCRVIRKVLSEKFCFMMWLGECFCLEPAE 489
Qy 481 GLVGDGHNDNEAYEGSEVDPAEPALHDVSGTYAVHGHQLEALYRALNVPODIAARASRLT 540
Db 490 GYVGDDGHNDNEAYEGSDVDPAESAIDISGTYVPGALOPLOALDPAEIAARACRLT 549
Qy 541 ATVELVASPDRLCEKRVLYGNKTRFTTVVDGAHLDEANGPEOYVLSFPAASROSGAGSHLT 600
Db 550 ATVELVASPDRLCEKRVLYGNKTRFTTVVDGAHLDEANGPEOYVLSFPAASROSGAGSHLT 609
Qy 601 YELTPGLQYRISNGIDCATFPFGAPSAAGGEVAFCSALYXRNRFQRISSLTGLW 660
Db 610 YASASGLEVRYAAGLDHAIIFAPGVSPSTGEVTAFCALYRNRRAQRISSLTGLW 669
Qy 661 LHPEGLIGLPPSPSGHIMESANPFCGEGTLYRTMS-TGCFSSDSPP--PEAAAPA 714
Db 670 FHPEGLIGLPPSPSGHIMESANPFCGEGTLYRTMS-TGCFSSDSPP--PEAAAPA 729
Qy 715 MAATPGLPHSTPPVSDIWLPPSPSEFQVDAHV-PPAPDPAGLPGFVLTTPPPPPVHK 773
Db 730 RAATPPL-----AALPPAPDPS--PEPFAPAPPEPAPGTT 764
Qy 774 PSIP-----PPSRNRLITYTPDGAKYVAGSLFSDCOMLVNANSPGRPGGICHAHYOR 829
Db 765 AGAPALITHOMARRRLFTYPDGSKVAFAGSLFSTCMTLVNANSPGRPGGICHAHYOR 824
Qy 830 FPEAFYPTFERIMEGLAAVTLTPRPIIHAAVAPDYVQNPRLAEAYRETCRSRGTAAVP 889
Db 825 YPASFDASVFMRDGAAYTLTPRPIIHAAVAPDYVQNPRLAEAYRETCRSRGTAAVP 884
Qy 890 LIGSGIYQVPSLSPDAMERNHRPGDELVLTPEPAAMFEANKPAQVLTITTEDTARTANL 949
Db 885 LIGSGIYQVPSLSPDAMERNHRPGDELVLTPEPAAMFEANKPAQVLTITTEDTARTANL 944
Qy 950 ALFIDATAEYGRACAGCTISPGIVHYOFTAGVGSKGSRSIOGDVYVVVPPRELNSM 1009
Db 945 ALFIDATAEYGRACAGCTISPGIVHYOFTAGVGSKGSRSIOGDVYVVVPPRELNSM 1004
Qy 1010 RRRGFAFTPTHTAARVYIGRRVYIDEAPSLPHLLILHMQASSVHLIDGPNQIPADFE 1069
Db 1005 RRRGFAFTPTHTAARVYIGRRVYIDEAPSLPHLLILHMQASSVHLIDGPNQIPADFE 1064
Qy 1070 HAGLVPAILRPELAPTSKMWYTHRCPADVCBLRGAVPKIOTSRVLSLFWNEPAIGOKL 1129
Db 1065 HAGLVPAILRPELAPTSKMWYTHRCPADVCBLRGAVPKIOTSRVLSLFWNEPAIGOKL 1124
Qy 1130 VTOQAKAANPAGATVEAGAGATFETTTATADARGLIOSSRAHVALTRHREKCVIL 1189
Db 1125 VTOQAKAANPAGATVEAGAGATFETTTATADARGLIOSSRAHVALTRHREKCVIL 1184
Qy 1190 DAPGLIREVGISDVYVNNPFLAGGEVGHRRPSVYPRGNPQUNGLTQAFPPSQISAHYQ 1249
Db 1185 DAPGLIREVGISDVYVNNPFLAGGEVGHRRPSVYPRGNPQUNGLTQAFPPSQISAHYQ 1244

```

QY	1250	LAELGHRPAPVAALVPCPELQGLLWYPOGLTSDSVLAFELTDIYHCBMAPSQRKA	13
QY	1250	LAELGHRPAPVAALVPCPELQGLLWYPOGLTSDSVLAFELTDIYHCBMAPSQRKA	13
Db	1245	LAELGHRPAPVAALVPCPELQGLLWYPOGLTSDSVLAFELTDIYHCBMAPSQRKA	13
QY	1310	VLSLTVGRYGRARRKLYEAASHDVRESLAFIPTIGPVQATTCCELYELVEAMVEKGDSGA	13
Db	1305	VLSLTVGRYGRARRKLYEAASHDVRESLAFIPTIGPVQATTCCELYELVEAMVEKGDSGA	13
QY	1370	VLELDLCNRDVSRIITFEFOKXCKNFETTGETIAHGKVGOGISAMSKTFCALFGPWRALKEK	14
Db	1365	VLELDLCNRDVSRIITFEFOKXCKNFETTGETIAHGKVGOGISAMSKTFCALFGPWRALKEK	14
QY	1430	ILALLPNTIIFYGAYEESYFAAASGACMFENDSEEPSTONNNSLGLCEYAMEEG	14
Db	1425	ILALLPNTIIFYGAYEESYFAAASGACMFENDSEEPSTONNNSLGLCEYAMEEG	14
QY	1490	MPQWLILYLHLVBSAMILLAPKESLKGFWKKGSGEPGLLNNTYWMNAITAHCEYRDPK	15
Db	1485	MPQWLILYLHLVBSAMILLAPKESLKGFWKKGSGEPGLLNNTYWMNAITAHCEYRDPK	15
QY	1550	VAAFGGDSVILCSDYRQSRNAALTAGCGTLKLYDYPRIGLYAGVYVAPGLTLPDVR	16
Db	1545	VAAFGGDSVILCSDYRQSRNAALTAGCGTLKLYDYPRIGLYAGVYVAPGLTLPDVR	16
QY	1610	FAGRLSEKMGPCPEAPBODRLAYCDELRLGNVAVGVDSVSVSGVGLVHNLIGL	16
Db	1605	FAGRLSEKMGPCPEAPBODRLAYCDELRLGNVAVGVDSVSVSGVGLVHNLIGL	16
QY	1670	QTADGKAHFTETIKRVLDTLNTSIQRYE	1698
Db	1665	QAVADGKAHFTESVREPLDTLNTSINCRVE	1693
RESULT	7		
Q089444	Q089444	PRELIMINARY:	PRF: 1693 AA.
AC	089444:		
DT	01-NOV-1996 (TEMBLrel. 01, Created)		
DT	01-NOV-1996 (TEMBLrel. 01, Last sequence update)		
DT	01-JUN-2000 (TEMBLrel. 14, Last annotation update)		
DE	UNNAMED PROTEIN PRODUCT.		
OS	Hepatitis E virus.		
OC	Viruses: ssRNA positive-strand viruses, no DNA stage.		
OX	NCBI_TaxID=12461;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RA	Yin S.R., Purcell R.H., Emerson S.U.:		
RL	Submitted (MAR-1996) to the EMBL/Genbank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=K52-87.		
RX	MEDLINE=93176571; PubMed=7871758;		
RA	Yin S., Purcell R.H., Emerson S.U.:		
RT	"A new Chinese isolate of hepatitis E virus: comparison with strains recovered from different geographical regions."		
RL	Virus Genes 9:23-32(1994).		
DR	EMBL: L25547; AAA91078.1; -		
DR	EMBL: L25595; AAA65468.1; -		
DR	INTERPRO: IPR000606; -		
DR	INTERPRO: IPR002588; -		
DR	INTERPRO: IPR002589; -		
DR	PFAM: PF01443; Viral_helicase1; 1.		
DR	PFAM: PF01660; Vmethyltransf. 1.		
DR	PFAM: PF01661; DUF271.1		
Q0	SEQUENCE 1693 AA; 185122 MW; 53914B8302338D5E CRC64;		

Query Match	82.7%;	Score 7456;	DB 14;	Length 1693;
Best Local Similarity	81.8%;	Pred. No. 0;		
Matches 1401;	Conservative 108;	Mismatches 161;	Indels 42;	Gaps 8;
OY	1	PGITTAIEQALAAANSALANNAVVRPPLSKQTELLINLMQPROLVPEPEVLMMHPRIOR	60	

Db	10	PGITAIJGQALIAANSALANAAVVRPFLSHQOIEILINLMQROCLVFRPEVFMHPIQR	69
Qy	61	VIHNELEOYCARGRCLEFGANPRSINDPNVLHRCFLRPVGRDYQRMYSAPITGPRAAN	120
Db	70	VIHNELEUYCARSGRCLEIGAHPRSINDPNVWHCFILRPARDQWRMYTAPITGPRAAN	129
Qy	121	GRSRLNCLRPADSTYCFEDGSRCAFAPENGVNLISHLMDIVDVAEAMARIGXTRLAA	180
Db	130	GRSRLNCLRPADSTYCFEDGSGCNFAETGLIALYLSLHDSPSDVAEAMFRIGMTRLAA	189
Qy	181	LHLPEVLLPBGTYHTTYSYLLIHDGDAVVYTBGDIISAGYNHDVSLIRAMIRTTKIVGDH	240
Db	190	LHLPEVLLPBGTYRASYLLIHDGRVVYTBGDISAGYNHDVSLIRAMIRTTKIVGTGH	249
Qy	241	PLVIERRAIIGCHVLLTLTAAPRSPMPRYPYTRRSTREVVYRSTIFGGGSPSLFPASCSTK	300
Db	250	PLVIERRAIIGCHVLLTLTAAPRSPMPRYPYTRRSTREVVYRSTIFGGGSPSLFPASCSTK	309
Qy	301	STFPAVNHIMDRMLTGLGAILDQAFCCSRILMYLIGISYKTVGALVANECSMNASDQAL	360
Db	310	STFPAVNAHIMDRMLTGLGAILDQAFCCSRILMYLIGISYKTVGALVANECSMNASDQAL	369
Qy	361	TAXITTAAYLTICHOYURLTQOAIKGMRLGVEHQAOKFTRLYSMLFEKSGRDYIDROLO	420
Db	370	TAVITTAAYLTICHOYURLTQOAIKGMRLEREHQAOKFTRLYSMLFEKSGRDYIDRGLOE	429
Qy	421	FYACCRMLWSGFLDPRVYLFPDESVPBCRTEFLKKVAKFCCPFMYLGOEETCFLGEAE	480
Db	430	FYACCRMLWSGFLDPRVYLFPDESAPCHCRTAIRAIVSKFCFPMYLGIOEETCFLGOAE	489
Qy	481	GLVGDHGHNDNEAYEGSEVDPABERANLIDVSGTAVHGHOLEALYRALNVPODIAARASRLT	540
Db	490	GVYDGDHGHNDNEAYEGSDVDAESAISDISGSSVVPSTALOPLYALDLPAEIVARAGRLT	549
Qy	541	ATVBLVASDRLCEBRTYLGKKTFRRTYVNGANILEANGRPOYLYLSDASQMSGASHSLT	600
Db	550	ATVAVSOVDGHIDEITLGNKKTFRTSYVDQAVLETNGBPRHNLSPDASQMSMAAPFSLT	609
Qy	601	YELRPAGLOVARISSNGLDCATFPPGAGAPSAARGEVAACSAALYVRNRTORHNSLTGSLM	660
Db	610	YASAGAEVRYVAAGLHDHRAVAPAGVPSRPSAGCEVTATCSALYRFRNRAQSLSLTGFMW	669
Qy	661	LHPGCLIGITPPSPSGHITWSANPFCGEGTLYTRTWS-----TSGFSSDQSPPE	709
Db	670	FHPRGLLGPRAPSPSGHITWSANPFCGEGSLYTRTWSADVASSPAQDGLFTSPSIPS	729
Qy	710	AAAAPAMAATGCLPSPSPSVSDIWLPPSPSEEPQVDAVRPAPR---DPAQLGCPVLLRP	766
Db	730	RAA---LTLTAAP-----LPPRAP---DPSLPPSAPRAGERA--PG---ATAG	766
Qy	767	PRPVHPRPSIPPSRNRRLLYTPDGAKVYAGSLFESDCDMLVNASNPGRNPRGGCLCHAF	826
Db	767	APALITHO-----TARHRRLLFTFPDCKSKYFAGSLFESTCTWLNVNASNDHNRGCGELCHAF	821
Qy	827	YQRPBEAFYPTFEMIRMGSLAAYTLPRPIIINAVAPRYRVEQONPKRYLEAARPTCSRGCTA	886
Db	822	YQRPASFSFDAQSEVMRQGAAYTLTLPRPIIINAVAPRYRLENNPKRYLEAARPTCSRLCTA	881
Qy	887	AYPLLGSGIYQVNVSLISFDAMERNHPRGDELYLTERPAMNPEANKPAQVLLTIEDTART	946
Db	882	AYPLLGIGIYQVNPISFDMERNHPRGDELYLTERAHPFANRPCTPILITIEDVART	941
Qy	947	ANLALIEDATBVGRRACAGCTISPGLYVHOFNAPGVSGSKSSIOGQGVDDVVVYPTRELR	1006
Db	942	ANLALIEDASTDVGRRACAGCRVTPGVYVOFTNAPGVSGSKSSITQADVDDVVVYPTRELR	1001
Qy	1007	NSMRRRGFAFTPHTAARVITGRVVIDEAPSLRPHLLILHMQRASSVHLLGDPQOIPAI	1066
Db	1002	NAMRRRGFAFTPHTAARVITGRVVIDEAPSLRPHLLILHMQRAATVHLLGDPQOIPAI	1061
Qy	1067	DPEHAGLVPAIRBELAPTSMWXYVTHRCADVCELIKGAVPKIOTTSRVLSLFWNERPAIG	1126


```

Db 1062 DEHAGLVPAIRBPLAPTSMMHVTNRCPADVCELLRGAPMIQIOTTSRVLSLFWGEPAVG 1121
Qy 1127 OKLVYQAQKANKANPGALITYHEAGATFTTITATADAGLIQSSAHAIVALTRTEKC 1186
Db 1122 OKLVYQAQKANKANPGSVTAEAGATYTTTITATADAGLIQSSAHAIVALTRTEKC 1181
Qy 1187 VIIDAPGLREVGISIVYNNFELAGGEVGHHPSPVIRPGNPONQITGOAFPSQOISA 1246
Db 1182 VIIDAPGLREVGISAIYNNFELAGGEVGHHPSPVIRPGNPONQITGOAFPSQOISA 1241
Qy 1247 YHQLAEELGHRPAPVAVALPCEPELEQGLLXMPDELTVSDVLYFELTDIVHCRMAPSQ 1306
Db 1242 FHLQAEELGHRPAPVAVALPCEPELEQGLLXMPDELTVSDVLYFELTDIVHCRMAPSQ 1301
Qy 1307 KRAVSTLVGRGRKTKLYEAASDVRESLARTPTIGVQATTGCTLYELVEAMVERKGD 1366
Db 1302 KRAVSTLVGRGRKTKLYEAASDVRESLARTPTIGVQATTGCTLYELVEAMVERKGD 1361
Qy 1367 GSAVLELDLCNDRSVITFEFOKKCNKFTTGETIAGKVGSGISAMSKTCALFGPWRAT 1426
Db 1362 GSAVLELDLCNDRSVITFEFOKKCNKFTTGETIAGKVGSGISAMSKTCALFGPWRAT 1421
Qy 1427 EKELIALLPNTIYGDAYEESVFAAASGAGSCMVEENDESEPDSTONNESLGLCEVME 1486
Db 1422 EKALIALLPQGYFYGDAPDDYFSAVAATAKASMWENDESEPDSTONNESLGLCEAIME 1481
Qy 1487 ECGMOMLRLYLHVASAVITLOAPKESLKGFKHSGEFTGLMNTVYMAVAITTAHCYER 1546
Db 1482 ECGMOMLRLYLHVASAVITLOAPKESLKGFKHSGEFTGLMNTVYMAVAITTHCYDR 1541
Qy 1547 DRRVAFAKGDSDSVILCSDRQSRNAALJAGGLKLVDRYRIGLVAGVVAAPGLTLP 1606
Db 1542 DLQVAFAKGDSDSVILCSDRQSRNAALJAGGLKLVDRYRIGLVAGVVAAPGLTLP 1601
Qy 1607 VVRFAGRLEKMMGPPEBAEQLRLAVCDFELRLTNVAOCVYDVSVRYGVSPGLVHNI 1666
Db 1602 VVRFAGRLEKMMGPPEBAEQLRLAVCDFELRLTNVAOCVYDVSVRYGVSPGLVHNI 1661
Qy 1667 GMLQTIADGKAHFTETIKPVLDITNSIIQRYE 1698
Db 1662 GMLQAVADGKAHFTESVKPVLDTLNTSILSRVE 1693

RESULT 8
ID 081876 PRELIMINARY: PRT: 1693 AA.
AC 081876:
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, last annotation update)
DE COMPLETE GENOME SEQUENCE.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=UIGH179;
RA Uchida T.;
RL Submitted (May-1992) to the EMBL/GenBank/DBJ databases.
DR EMBL; D11093; BAA01865.1; -
DR INTERPRO: IPR000606; -
DR INTERPRO: IPR002589; -
DR INTERPRO: IPR002588; -
DR PFM; PF01443; Viral_helicase1.1.
DR PFM; PF01600; Vmethyltransf.1.
DR PFM; PF01661; DUF27.1.
SQ SEQUENCE 1693 AA; 184839 MW; 4E4222ED55AE0B1 CRC64;

```

Query Match 82.4%; Score 7424; DB 14; Length 1693;
 Best Local Similarity 81.6%; Pred. No. 0;
 Matches 1397; Conservative 108; Mismatches 165; Indels 42; Gaps 8;

```

Qy 1 PGITTAIEQALAAANSALANAVVPRFLSRVOTELINLMOROLVPRREVLMNHPOR 60
Db 10 PGITTAIEQALAAANSALANAVVPRFLSHQOIELINLMOROLVPRREVLMNHPOR 69
Qy 61 VINELEOYCARAGRCLEVGAPRSINDPNVLRHCFRLPYRVDQVQWYSAFTRGPAN 120
Db 70 VINELEOYCARAGRCLEVGAPRSINDPNVLRHCFRLPYRVDQVQWYSAFTRGPAN 129
Qy 121 CRSSALRGLEPADRYCFDFGSCAPAEETGVALYSLHDIIMPADVAEMARHGSTRLYAA 180
Db 130 CRSSALRGLEPADRYCFDFGSCAPAEETGVALYSLHDIIMPADVAEMARHGSTRLYAA 189
Qy 181 LHLPEVLLPGTYHTSYLLHDSRAVTVTGDPSAGCNHNVSTLRARITRTKIVGDH 240
Db 190 LHLPEVLLPGTYHTSYLLHDSRAVTVTGDPSAGCNHNVSTLRARITRTKIVGDH 249
Qy 241 PLVIEVRATGCHFFVLLTAAPRSPMPVYPRSTEVYRSIFGSGSPLEPSACSTK 300
Db 250 PLVIEVRATGCHFFVLLTAAPRSPMPVYPRSTEVYRSIFGSGSPLEPSACSTK 309
Qy 301 STFHAVPVHIMDLMPGATLDDQAFCCSLMTYLRNGISYKTVGALVANEKNASSEDAL 360
Db 310 STFHAVPVHIMDLMPGATLDDQAFCCSLMTYLRNGISYKTVGALVANEKNASSEDAL 369
Qy 361 TAXITAAVLTICHORYLRTQALISGMRLGVEHAOKFITLTYSMLEKSGROVTPGROL 420
Db 370 TAXITAAVLTICHORYLRTQALISGMRLGVEHAOKFITLTYSMLEKSGROVTPGROL 429
Qy 421 FYAQCRRLWSAGFHLDPRLVLPFDESVCRCRTLTKAKAGFCGFMRLGCECTFLEPAE 480
Db 430 FYAQCRRLWSAGFHLDPRLVLPFDESVCRCRTLTKAKAGFCGFMRLGCECTFLEPAE 489
Qy 481 GLVGDGHNDNEAVEGSEVDPBAERHLDVSGTVAVHGHOLEALYRALNPODIAARSLT 540
Db 490 GLVGDGHNDNEAVEGSEVDPBAERHLDVSGTVAVHGHOLEALYRALNPODIAARSLT 549
Qy 541 ATVELVASPRLCRFLYLGKTRTYVVDGAHLNANGPEQVYLSFASRQSMGAGSHLT 600
Db 550 ATVELVASPRLCRFLYLGKTRTYVVDGAHLNANGPEQVYLSFASRQSMGAGSHLT 609
Qy 601 YELTPAGLOVRISNGIDCTATFPPGAPSAABGEVAFCALYRYNRFOQRSLTGLTW 660
Db 610 YELTPAGLOVRISNGIDCTATFPPGAPSAABGEVAFCALYRYNRFOQRSLTGLTW 669
Qy 661 LHPGGLIGTPPSPGHIIMSANPFCGEGTLTYRTMS-----TSGFSDSPPE 709
Db 670 LHPGGLIGTPPSPGHIIMSANPFCGEGTLTYRTMS-----TSGFSDSPPE 729
Qy 710 AAPAPMAATPGLHSPRPVSDIVLPPPSSEFQVDAAPVPAP---DPAGLPGVVLTLP 766
Db 730 RAA---LTPAAP-----LPPAP---DPSPTSAARGPAP---ATAG 766
Qy 767 PPPVHKPSIPPSRNRLTYTPDGAKVYAGSLFESDCMLVNASNPGRPPGGGLCHAF 826
Db 767 AAPATHQ-----TARHNRLLFTYPDGSKVYAGSLFESTCTWLVNASNVDRPPGGGLCHAF 821
Qy 827 YQFPEAFYPTERIMEGLAAYTLTPRPIIHAAVPDIYVQNPRLKLEAAYRETCSRRTGA 886
Db 822 YQFPEAFYPTERIMEGLAAYTLTPRPIIHAAVPDIYVQNPRLKLEAAYRETCSRRTGA 881
Qy 887 AYPFLSGIYQVYVLSFQWERNHRGDELXLTLEPAANFEANKRAQVLTTEEDART 946
Db 882 AYPFLSGIYQVYVLSFQWERNHRGDELXLTLEPAANFEANKRAQVLTTEEDART 941
Qy 947 ANALEIDATEVGRACAGCTISPGIYHYQFTAGVGSGRSIRIOQDDVVVVVPTRELR 1006
Db 942 ANALEIDATEVGRACAGCTISPGIYHYQFTAGVGSGRSIRIOQDDVVVVVPTRELR 1001
Qy 1007 NSWRRRGFAAFTPTAARTYIGRRVVIDEAPSLPPLLHLMORASSVHLGDPNDIPAT 1066
Db 1002 NAWRRRGFAAFTPTAARTYIGRRVVIDEAPSLPPLLHLMORASSVHLGDPNDIPAT 1061
Qy 1067 DEHAGLVPAIRBPLAPTSMMHVTNRCPADVCELLRGAPMIQIOTTSRVLSLFWNEPAIG 1126

```



```

|||||
Db 1062 DFEHAGLVPAIRPDLAPTSMMHWHTRCPADVCCELRCAGYPMIQTTSRYLSRSLFWMEPAVG 1121
QY 1127 OKIYXTOAKANAGATVYHBAOGATFETETIATADARGLIOSSRAHAYALPHTEKCC 1186
Db 1122 OKIYXTOAKANAGATVYHBAOGATFETETIATADARGLIOSSRAHAYALPHTEKCC 1181
QY 1187 VILDPAGLIREVIGSDVYVNNFPLAGSEVGHRSVPVIRGNDDOGLTQAFPPSCQISA 1246
Db 1182 VITDAPGLVREVATSDALVNNFPLAGSEIGHORSPVICGNPDANVDLAPFPSCQISA 1241
QY 1247 YHOLAEELGHRPAPYAAVLPCCPELEGLYMPQELVSDSVLVELFDIYHCRMAPSQ 1306
Db 1242 FHOALAEELGHRPAPYAAVLPCCPELEGLYLPQELTTCDSVTFELDIYHCRMAPSQ 1301
QY 1307 RKAVALSTLVGRGRTKLYEAAHSDVRESLARFPTIGPVQATCELEIYEAWEKGD 1366
Db 1302 RKAVALSTLVGRGRTKLYEAAHSDVRESLARFPTIGPVQATCELEIYEAWEKGD 1361
QY 1367 GSAVLELDLCNRDYSRTTFEQKXCNKFTTGETIAHGKVGOGISAMSKTFCALFGWFRAT 1426
Db 1362 GSAVLELDLCNRDYSRTTFEQKXCNKFTTGETIAHGKVGOGISAMSKTFCALFGWFRAT 1421
QY 1427 EKELLALPRIFEGDAEESVFAAASGASCVFENDESEPOSTONNESLGECEVME 1486
Db 1422 EKALLALPGCVFEGDADDTVFSAVAASAVFENDESEPOSTONNESLGECEVME 1481
QY 1487 ECGMPDWLIRLYHVRSAWILQAPKESLKGFKHSGEPGLTNNTVNNMAIIAHCEFR 1546
Db 1482 ECGMPDWLIRLYHVRSAWILQAPKESLKGFKHSGEPGLTNNTVNNMAIIAHCEFR 1541
QY 1547 DFRVAARFGDDSVVLCSDYROSNNAAIIAGCGKLKLVDPYRIGLYAGVVAAPGLGLPD 1606
Db 1542 DLQVAARFGDDSVVLCSDYROSNNAAIIAGCGKLKLVDPYRIGLYAGVVAAPGLGLPD 1601
QY 1607 VVRVAGRLSEKMGKGPRAQOLRLAVCDLRLTNVNOGVVYVSRVYSGVSPGLVHMLI 1666
Db 1602 VVRVAGRLSEKMGKGPRAQOLRLAVCDLRLTNVNOGVVYVSRVYSGVSPGLVHMLI 1661
QY 1667 GMLQTIADGKAHFTETIKPVLDTNLSIIQRYE 1698
Db 1662 GMLQAVADGKAHFTESVAPVLDLNSILCRVE 1693

RESULT 9
Q81862 PRELIMINARY; PRT: 1693 AA.
AC Q81862:
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE ORF 1.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.C., Bradley D.W., Fry K.E.,
RA Reyes G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
RT full-length viral genome.";
RL Virology 185:120-131(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92261377; PubMed=1584074;
RA Uchida T., Suzuki K., Hayashi N., Iida F., Hara T., Oo S.S.,
RA Wang C.K., Shikata T., Ichikawa M., Rikihisa T., Mizuno K., Wain K.M.;
RT "Hepatitis E virus: cDNA cloning and expression.";
RL Microbiol. Immunol. 36:67-79(1992).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92335008; PubMed=1630924;
RA Aye T.T., Uchida T., Ma X.Z., Iida F., Shikata T., Zhuang H.,
RA Wain K.M.;
RT "Complete nucleotide sequence of a hepatitis E virus isolated from the
RT Xinjiang epidemic (1986-1988) of China.";
RL Nucleic Acids Res. 20:3512-3512(1992).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=92115700; PubMed=1731327;
RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,
RA Malik I.A., Iqbal M., Purcell R.H.;
RT "Characterization of a prototype strain of hepatitis E virus.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=92271462; PubMed=1589964;
RA Fry K.E., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,
RA Platak M., Feldman R.A., Yun K.Y., Purdy M.A., Et AL.;
RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
RT region encoding consensus motifs for an RNA-dependent RNA polymerase
RT and an ATP/GTP binding site.";
RL Virus Genes 6:173-185(1992).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=93348763; PubMed=8346669;
RA Bi S.U., Purdy M.A., McCausland K.A., Margolis H.S., Bradley D.W.;
RT "The sequence of hepatitis E virus isolated directly from a single
RT source during an outbreak in China.";
RL Virus Res. 28:233-247(1993).
DR EMBL: L08816; AAA03189.1;
DR INTERPRO: IPR000606;
DR INTERPRO: IPR002588;
DR INTERPRO: IPR002589;
DR PFAM: PF01443; Viral_helicase1;
DR PFAM: PF01660; Vmethyltransf. 1;
DR PFAM: PF01661; DUF27; 1;
SQ SEQUENCE 1693 AA; 185052 MW; D3363602435B8FF5 CRC64;

Query Match 82.2%; Score 7412; DB 14; Length 1693;
Best Local Similarity 81.7%; Pred. No. 0;
Matches 1395; Conservative 106; Mismatches 174; Indels 32; Gaps 7;

QY 1 PGTTAIEQALAAANSALANAVVVPFLSRVQTEILINLMOPQOLVFRPVLLNHPQR 60
Db 10 PGTTAIEQALAAANSALANAVVVPFLSRVQTEILINLMOPQOLVFRPVLLNHPQR 69
QY 61 VIHNELEOYCARAGRCLEVGAPHRSTINDNPNVLRHCFRLPVGDYORWYSAPRGPAN 120
Db 70 VIHNELEOYCARAGRCLEVGAPHRSTINDNPNVLRHCFRLPVGDYORWYSAPRGPAN 129
QY 121 CRRSALRGLEPPADRTYCFDGFSCAFPAETGVALYSLHDLMPADVAEAMARHGXTRLVAA 180
Db 130 CRRSALRGLEPPADRTYCFDGFSCAFPAETGVALYSLHDLMPADVAEAMARHGXTRLVAA 189
QY 181 LHPPEVLLPPGTYHTTSYLLIHGDGRVAVYVEEDTSAGYVHNDVSLRAMVTRTKIYGDH 240
Db 190 LHPPEVLLPPGTYHTTSYLLIHGDGRVAVYVEEDTSAGYVHNDVSLRAMVTRTKIYGDH 249
QY 241 PLVTERPAICGCVLLTLTAPEPSMPYVPYPRSTEVYVNSIGPGSPSLFSPASCT 300
Db 250 PLVTERPAICGCVLLTLTAPEPSMPYVPYPRSTEVYVNSIGPGSPSLFSPASCT 309
QY 301 STEFAVPHVHIDRLMFGATLDQAFCCSRMLTYLRGISYKVTVGALVANAGNASEDAL 360
Db 310 STEFAVPHVHIDRLMFGATLDQAFCCSRMLTYLRGISYKVTVGALVANAGNASEDAL 369
QY 361 TAXTAAYLTICHOYRLRTOAISGMRRLGVENAKITRLIYSLFEKSGDYIPGROL 420
Db 370 TAXTAAYLTICHOYRLRTOAISGMRRLGVENAKITRLIYSLFEKSGDYIPGROL 429
QY 421 FYACCRMLSGFHLDRPVLVFDESVPCCRTFLKVAAGKCCPMRMIGOCCTFLPAE 480
Db 430 FYACCRMLSGFHLDRPVLVFDESVPCCRTFLKVAAGKCCPMRMIGOCCTFLPAE 489

```



```

Db 430 FYACRRRLSGFHLDPVLYVDESAPCHCTVIRKALSKFCCKMKGJQECTCLQPAE 489
QY 481 GLVGDHGDNEAYEGSEVDPAEPALHDVSGYVAHGHOLEALRYALANVODIAAASRLT 540
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 490 GVVGDGHDNDSYEGSDVDPAESALSDISGSYYVGTALQPLQYQALDDELVAACHLT 549
QY 541 ATVELVASPDLRECTVYCNKTFRTVVDCALHEANGPEQYVLSFSDASROSKAGSHSLT 600
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 550 ATYKASQYDGRIDETLLGNKTFRTSFVDGAVLETNGPERNHLSPDASQSMAGPFSILT 609
QY 601 YELPAGLQVITSSNGLDCTATFPFGCAPSAAPGEVAAGCSALRYNFEPTQSHSLTGLM 660
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 610 YAAASAGLEVYVAGAGDLHRAIFAPGVPSPRNGEVITAFCSALIKFNREQHSLTGLM 669
QY 661 LHPGLGIFPPFPBGHIMESANPECCGTLTYRTWS-----TSGFSSDFSPPE 709
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 670 FHPGLGIFAPFSPBGHIMESAKPCCGTLTYRTWS-EVDAVSSPTRDLFMSE--PP- 726
QY 710 AAAAAMATPCLPHSTPVPVSDIWLPPPESEFOVDAAPVPAPDPAGLPGEVYLTTPPPE 769
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 727 --TFSRAATPPL-----AAFLP-----LAPDPS 749
QY 770 PVHRPSIIP-----SNRRLLTYPDGAKYVAGSLPESDCDWLVNANP 814
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 750 PSSAPALDEPASAATSGVPALTHTOTAKRRLFTYPDGSKYFAGSLSESTLWLNANSV 809
QY 815 GHRPGGGLCHAFYORPEAFYPTFEIMREGLAAYTLPRPIIHAVDPDYVEONPKRLA 874
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 810 DHCGGGGLCHAFYORPASFDAFCVMDGAAYTLTPRPIIHRAPDYRLHNPKRLEA 869
QY 875 AYRRTCSRGTAAPPLGSGIYQVPSISFLAMERNHNPGEYLITPEANWFEANKPAQ 934
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 870 AYRRTCSGLTAAPPLGTGIYQVIGISFDMERNHNPGEYLITPEANWFEANKPETR 929
QY 935 PVLITEDTARTANLALTEIDATEVGRACACTISPGIVHQFTAGVSGSKSSIOGD 994
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 930 PVLITEDTARTANLALTEIDATEVGRACACTISPGIVHQFTAGVSGSKSSITRAD 989
QY 995 VDVVVVPTRELNSWRRRGFAAFTPHTAARVYIGRRVVIDEAPSLPHLLHMQRASSV 1054
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 990 VDVVVVPTRELNSWRRRGFAAFTPHTAARVYIGRRVVIDEAPSLPHLLHMQRATV 1049
QY 1055 HULGDPNPIAIDFEBHGLVVAIRPELAPTSWXYTHRCPADVCELLRGAVPKIQTTSRV 1114
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1050 HULGDPNPIAIDFEBHGLVVAIRPELAPTSWXYTHRCPADVCELLRGAVPMIQTTSRV 1109
QY 1115 LRSLEWNEPAIGOKLYVTOAKAANPGAITVHEAGATFETTLITADAGLIQSSRAH 1174
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1110 LRSLEWNEPAIGOKLYVTOAKAANPGAITVHEAGATFETTLITADAGLIQSSRAH 1169
QY 1175 AIVALTHTTEKCVILDPGLLREVGISDVIYNNEFLAGGEVGHRRPSVTPRGNDONLGT 1234
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1170 AIVALTHTTEKCVILDPGLLREVGISDVIYNNEFLAGGEVGHRRPSVTPRGNDONLGT 1229
QY 1235 LQAPPPSCQISAHYQALAEELGHRAPVAAYLPCCPELEQGLLIMPOLITYSDSVLEVLT 1294
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1230 LQAPPPSCQISAHYQALAEELGHRAPVAAYLPCCPELEQGLLIMPOLITYSDSVLEVLT 1289
QY 1295 DIYHCRMAASORKAAYSTVIGRGRTKLYEAHSDVRESLARTFTIGVQATTCLELY 1354
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1290 DIYHCRMAASORKAAYSTVIGRGRTKLYEAHSDVRESLARTFTIGVQATTCLELY 1349
QY 1355 ELVEANVEKODGSVAVLELDCNRDVSRIFFOKCKNFTTGETIAGKVGOGISANSKT 1414
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1350 ELVEANVEKODGSVAVLELDCNRDVSRIFFOKCKNFTTGETIAGKVGOGISANSKT 1409
QY 1415 FCALFGFWPRAIEKELIALLPNIIFYDAYEESVYAAVAGSAGSCWVFENDSEFSDTON 1474
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1410 FCALFGFWPRAIEKELIALLPNIIFYDAYEESVYAAVAGSAGSCWVFENDSEFSDTON 1469
QY 1475 NESTLGLCEVVMESCGMOWILRLYLHVSAMIIQAPKESLKGPKKKSGEPGILLMNTVW 1534
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1470 NESTLGLCEVVMESCGMOWILRLYLHVSAMIIQAPKESLKGPKKKSGEPGILLMNTVW 1529

```

```

QY 1535 NMAIIACHYERDFEVAFAFKDDSVYLCSDYRQSRNAAALIAGGCLKLVDPYRIGLYAG 1594
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1530 NMAVITHCYDRDLQVAFKDDSDIVLCSEYRQSGAAVILAGGCLKLVDPYRIRLYAG 1589
QY 1595 VVAAGLCTLPDVVRFAGRLSEKMWGPGPERAEQRLIACVDFLGLTNVAQVCDVVSrv 1654
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1590 VVAAPGLCALPDVVRFAGLREKMWGPGPERADELRIVASDPLRKTITVNAQVCDVVSrv 1649
QY 1655 YGVSGLVHNLIGMLQTLADGKAHFTETIRKPVLLTNSIIORVE 1698
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1650 YGVSGLVHNLIGMLQTLADGKAHFTETIRKPVLLTNSIIORVE 1693

RESULT 11
039221
ID 039221 PRELIMINARY; PRT: 1693 AA.
AC 039221;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE NONSTRUCTURAL POLYPEPTIDEIN.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HYDERABAD, INDIA;
RA Asari I.H., Nanda S.K., Durapal H., Jameel S., Panda S.K.;
RT "Eukaryotic expression of nonstructural protein (NSP1) of HEV: absence
RT of any processing.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF028091; AAB82002.2; -
DR INTERPRO: IPR006066; -
DR INTERPRO: IPR002588; -
DR INTERPRO: IPR002589; -
DR PIRAM: PF01443; Viral_helicase1; 1.
DR PIRAM: PF01660; Vmehylitransf.; 1.
DR PIRAM: PF01661; DUF27; 1.
KW Polypeptidein.
SQ
SEQUENCE 1693 AA: 185562 MW: 7AFED003532E12BC3 CXC64:

Query Match 81.3%; Score 7329; DB 14; Length 1693;
Best Local Similarity 79.7%; Pred. No. 0;
Matches 1374; Conservative 121; Mismatches 163; Indels 66; Gaps 6;
QY 1 PGITTAIEQAAALAAANSALANAVVVRPPLSRVQTELLINLMQRPOLVFRPEVTANHPIDR 60
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 10 PGITTAIEQAAALAAANSALANAVVVRPPLSRVQTELLINLMQRPOLVFRPEVTANHPIDR 69
QY 61 VINHELYOYCARARGLCEVGAHPRSTINDPNVYHRCFLRPVGDVQRYSAIPRGPAAN 120
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 70 VINHELYOYCARARGLCEVGAHPRSTINDPNVYHRCFLRPVGDVQRYSAIPRGPAAN 129
QY 121 CRSSALRGLPADRTYCFDGSFRCAPFAETGVALYSLHDLMPADVAAMAHGXTRLYAA 180
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 130 CRSSALRGLPADRTYCFDGSFRCAPFAETGVALYSLHDLMPADVAAMAHGXTRLYAA 189
QY 181 LHPPEVLLPPTYHTTSYLLIHDGDRVVYVSGDTSAGYNHDSVILRAMIRTKIYGDH 240
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 190 LHPPEVLLPPTYHTTSYLLIHDGDRVVYVSGDTSAGYNHDSVILRAMIRTKIYGDH 249
QY 241 PLVIERRAIGCHFVILLTAPAPSPRPVYPRSTVYVRSIFGPGSGSLSPSACSTK 300
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 250 PLVIERRAIGCHFVILLTAPAPSPRPVYPRSTVYVRSIFGPGSGSLSPSACSTK 309
QY 301 STFHAHVHIMDRMLFEGATLDQAFCCSRMLTYLRGISYKVVGALVYANEGNNASEDAL 360
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 310 STFHAHVHIMDRMLFEGATLDQAFCCSRMLTYLRGISYKVVGALVYANEGNNASEDAL 369
QY 361 TAXITAAVLTICQORYLTCQAIKGMRLGVEHAQKFTTLYLWMLFEKSGRDYIPGROLQ 420
    |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

```

Db 370 TAVITAAVLTICHOVYLRTOAISKIRLEREHQKFTIRLXSWLFEKSGRDYIPGRQLE 429
Qy 421 FTAQCRRLISAGFHLDPRLVYDESVPCRCRTFLKLVAGKFFCCFMRILQOECTCELEPE 480
Db 430 FTAQCRRLISAGFHLDPRLVYDESVPCRCRTFLKLVAGKFFCCFMRILQOECTCELEPE 489
Qy 481 GLVGHGHNEVYEGSEVDPAEPALDVSGTVAVGHOLEALYRLANVPODIAAARSRLT 540
Db 490 GVVAGQGHNEVYEGSEVDPAEPALDVSGTVAVGHOLEALYRLANVPODIAAARSRLT 549
Qy 541 ATVELVAPDLRECTVLGNKTFRTTVVDGAHLLEANGPEOYVLSFDASROSAGASHSLT 600
Db 550 ATVKXSYQVDRIDCETILGNKTFRTSPYDGAVLETNNGPRNHLSTPDASOSTMAAGFSLT 609
Qy 601 YELPAGLQVRISSNGLDCTATPPGAGSAPAEVAAFCSLXRYNRFTQRHSLTGLIM 660
Db 610 YAAASAGLEVRYGAGLDHRAIFAPGVSPRNPGEVTAFCALYRFRNRAQHSLSLTGNIM 669
Qy 661 LHPEGLLIFPPFSGHITMESANPFCGEGTLTTRTWS-----TSGFSSDPSPPE 709
Db 670 FHPEGLIGLFAFPFSGHWSAKPFCGEGTLTTRTWS-----TSGFSSDPSPPE 726
Qy 710 AAAPMAATPGDLPSTPVDIWLPPSEEFQVDAAPVPAAPDAGLPGVVLTPPPPP 769
Db 727 --IPSRATPPL-----AAAPLP-----LAPDSP 749
Qy 770 PVHKPSIPP-----SRNRLLYTPDGAKVYAGSLFESDCMLVANSP 814
Db 750 PSSAPALDEPASAAVNSGVAITHQTAHRRLLETFYPDGSKVPAGSLFESTCTWLVANSV 809
Qy 815 GHRPGGICHAHYORPREAFPTFEFMRGLAAYTLTPRIIHAAPDVRYQNPRLRLA 874
Db 810 DHCPGGICHAHYORPREAFPTFEFMRGLAAYTLTPRIIHAAPDVRYQNPRLRLA 869
Qy 875 AYRETCRSRGTAAYPLDGSIVQVNSLSDAMERNHRGDELYLEPAANFEANRPAQ 934
Db 870 AYRETCRSRGTAAYPLDGSIVQVNSLSDAMERNHRGDELYLEPAANFEANRPAQ 929
Qy 935 PYLTTTETDARLANALEIDAATENVGRACAGTISPGIYHYOFTAGVPGSGKSRSITQGD 994
Db 930 PYLTTTETDARLANALEIDAATENVGRACAGTISPGIYHYOFTAGVPGSGKSRSITRAD 989
Qy 995 VVVVVVPRRELNSMRGFAFTPTARVITGRVVYIDEAPSLRPHLLHMQAASSV 1054
Db 990 VVVVVVPRRELNSMRGFAFTPTARVITGRVVYIDEAPSLRPHLLHMQAASSV 1049
Qy 1055 HLLGDPNQIPAIIDFEGHAGLVPAIRPELAPTSMWXYTHRCADVCELIRGAYPKIQTTSRV 1114
Db 1050 HLLGDPNQIPAIIDFEGHAGLVPAIRPELAPTSMWXYTHRCADVCELIRGAYPKIQTTSRV 1109
Qy 1115 LRSLEWNEPAIQOKLYXTOAKAANGAIVTHEAGATETETIATADARGLIOSSRAH 1174
Db 1110 LRSLEWNEPAIQOKLYXTOAKAANGAIVTHEAGATETETIATADARGLIOSSRAH 1169
Qy 1175 AIVALTFRHTEKVIIDAPGLREVGISDVIYNNFELAGGEIGHQRSVIRPGADNAVPT 1234
Db 1170 AIVALTFRHTEKVIIDAPGLREVGISDVIYNNFELAGGEIGHQRSVIRPGADNAVPT 1229
Qy 1235 LOAFPPSCQISAYHOLAEELGHRPAVAVLPPCPLEGLYMPOLTVSDSVLEFLT 1294
Db 1230 LOAFPPSCQISAYHOLAEELGHRPAVAVLPPCPLEGLYMPOLTVSDSVLEFLT 1289
Qy 1295 DIVHCRMAAPSORKAVLSTLVGRYGRRTKLYEASHSDVRESLARFTPTGTGPOVQATTCELY 1354
Db 1290 DIVHCRMAAPSORKAVLSTLVGRYGRRTKLYEASHSDVRESLARFTPTGTGPOVQATTCELY 1349
Qy 1355 ELVEAMVEKGOGSANVELDLCNRDYSRTTFPOKXCNKFTTGETTAHGAVGQISMSMT 1414
Db 1350 ELVEAMVEKGOGSANVELDLCNRDYSRTTFPOKXCNKFTTGETTAHGAVGQISMSMT 1409
Qy 1415 FCALFGPWRRAIEKELLALPNIFGDAYEESVFAAASGAGSCAVFENDESEFSTON 1474
Db 1410 FCALFGPWRRAIEKELLALPNIFGDAYEESVFAAASGAGSCAVFENDESEFSTON 1469

```

```

Qy 1475 NFSLIGECVMECCGMPQMLIRLYHVRSAWTLQAKRESLKGWKKHSGEPGLTMNTV 1534
Db 1470 NFSLIGECVMECCGMPQMLIRLYHVRSAWTLQAKRESLKGWKKHSGEPGLTMNTV 1529
Qy 1535 NMAIITHACEYFDFRYAAAKGDDSVYLCSDYROSNAALIGCCGKLKVDYPIGLYAG 1594
Db 1530 NMAIITHACEYFDFRYAAAKGDDSVYLCSDYROSNAALIGCCGKLKVDYPIGLYAG 1589
Qy 1595 VVAVPGLTLPDVVRPAGRLSEKNMGPGPERAEQRLAYCDFLRLTNVAQVCVDVYSRV 1654
Db 1590 VVAVPGLTLPDVVRPAGRLSEKNMGPGPERAEQRLAYCDFLRLTNVAQVCVDVYSRV 1649
Qy 1655 YGVSPGLVNLIGMLQITADGKAHFTETIKPYLDTNLSITQRYE 1698
Db 1650 YGVSPGLVNLIGMLQITADGKAHFTETIKPYLDTNLSITQRYE 1693

RESULT 12
Q81344 PRELIMINARY; PRT; 1693 AA.
AC Q81344;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, last sequence update)
DE 01-JUN-2000 (Tremblrel. 14, last annotation update)
DE HEPATITIS E VIRUS COMPLETE GENOME.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
RN NCBI_TaxID=12461;
RX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HEBEL;
RX MEDLINE=93348763; PubMed=8346669;
RA Bi S.L., Purdy M.A., McCausland K.A., Margolis H.S., Bradley D.W.;
RT "The sequence of hepatitis E virus isolated directly from a single
RT source during an outbreak in China."
RL Virus Res. 28:233-247(1993).
DR EMBL: M94177; AAA96139.1; -.
DR INTERPRO: IPR000606; -.
DR INTERPRO: IPR002588; -.
DR INTERPRO: IPR002589; -.
DR PFAM: PF01443; Viral_helicase1; 1.
DR PFAM: PF01660; Vmethyltransf; 1.
DR PFAM: PF01661; DUF277; 1.
SQ SEQUENCE 1693 AA; 185271 MW; E3308F128E0B9C6F CRC64;

Query Match 81.0%; Score 7304; DB 14; Length 1693;
Best Local Similarity 80.8%; Pred. No. 0;
Matches 1380; Conservative 103; Mismatches 192; Indels 32; Gaps 7;

Qy 1 PGTTAIEQAALAAANSALANAVVPRPLSRVOTELINLMOPROLVFREVLMNHPOR 60
Db 10 PGTTAIEQAALAAANSALANAVVPRPLSRVOTELINLMOPROLVFREVLMNHPOR 69
Qy 61 VTHNELCYCRARAGCLEYGAPHSINDPNVYLHCFLRPVGRDYQKWTSAFTRGPAN 120
Db 70 VTHNELCYCRARAGCLEYGAPHSINDPNVYLHCFLRPVGRDYQKWTSAFTRGPAN 129
Qy 121 CRRSALRGLPAPARVYCFQFSGFSCAFAEFGVALYSLHDMPDVADVAMARHGHXTLYXA 180
Db 130 CRRSALRGLPAPARVYCFQFSGFSCAFAEFGVALYSLHDMPDVADVAMARHGHXTLYXA 189
Qy 181 LHLPRVLLPPGTYHTTSTYLLIHDGDAVVTYBGDSAGYNHDVSLIRAWIRTKTIVGDH 240
Db 190 LHLPRVLLPPGTYHTTSTYLLIHDGDAVVTYBGDSAGYNHDVSLIRAWIRTKTIVGDH 249
Qy 241 PLVIEVRALGCFVULLTAAPRSPMPVYPRRSTREYVYKSTFGGGSPLFPSCSTK 300
Db 250 PLVIEVRALGCFVULLTAAPRSPMPVYPRRSTREYVYKSTFGGGSPLFPSCSTK 309
Qy 301 STPHAVVHMDLMLFGATLDDQAFCCSLMYLGLGISKVVGVGLVYNEGNASBDAL 360
Db 301 STPHAVVHMDLMLFGATLDDQAFCCSLMYLGLGISKVVGVGLVYNEGNASBDAL 360

```

D	b	310	RPHNAVPHAMIDRLMELGATLDDQAFCCSRLMITYLNGISYKVTYCVGLVANBGMNASEVAL	369
Q	y	361	TAXITAAVLITCHORYLRTQTAISKMRRLGVEHAQKFITRLYSWLEKSGRDYIPGRQLQ	420
D	b	370	TAVITAAVLITCHORYLRTQTAISKMRRLREHAQKFITRLYSWLEKSGRDYIPGRQLQ	429
Q	y	421	FYACRRMLSGFHLDRVLYPDESVYQCRRTFLKKVYKAPCCPMYMLGOEOTCGLPEAE	480
D	b	430	FYACRRMLSGFHLDRVLYPDESBAPCHRTAIRKAVSFCCPMYMLGOEOTCGLPEAE	489
Q	y	481	GLVGDHGHDEAGYSGSEVDPAEPAHLDVSGTVAHVHOLEEALYRALNMPODITAAASRLT	540
D	b	490	GAVGDQGHNDNAYGSDVDPDAESALDSISGTYVPETALQPLYLQALDPAELVARAGSLT	549
Q	y	541	ATVELVASPDLRECRYLCKTFRITTYVDCAHLEANGPEQYVLSFDSASQSGACSHSLT	600
D	b	550	ATVAVSOVDKIDDELTLGNKKTERTSFGVAVLEITNGPERHNLSPFASQSTMAAPFSLT	609
Q	y	601	YELPAGIAYRISNGIDCATFPPGAPAAAGCEVAACSLARYNRPTQHSILTLGLGM	660
D	b	610	YASASAGLEVRYAAGJGDHRAVAPGVSPPASAGCEYATCASLYRNRAPQLSLTGNFW	669
Q	y	661	LHPGELGIFPPSPGCHIMESANPFCGEGTLYRTWS-TSGFSSPSP-----PEAAPA	714
D	b	670	FHPGELGIFAPFSPBGHVMESANPFCGEGTLYRTWSSEVDANSSPAPDPLGTSISPSTPS	729
Q	y	715	MAATPGLPHSTPPVSDIWLVPPESEFQVDAVAP-PPADPAGLP-GRVVLTPPPPPVH	772
D	b	730	RAAT-----LTP-----AAPLPAPDPPSPSPASAPGAPGATAR	766
Q	y	773	KPSLP-PESNRRLLYPPGCAKYVAGSLPESQDMLVANSNPHRRPGGGLCHATYQFPR	831
D	b	767	APALTHQARRRLLEFYPGGSYKVPAGSLPESQCYLVANSVNDHRRPGGGLCHATYQFPR	826
Q	y	832	EAFPTFERIMREGILAAATLPPRLIHAVAADYVEQNPRLAAARETCSRGTAAAYPL	891
D	b	827	ASFLPASFSVMDGGAATLTPRPIHAHAVADYLEINPMLAEARRETCSRGTAAAYPL	886
Q	y	892	GSGIYQVPSLSFDAMERNHRPDELYLTERPAANMEANKPAQVLTITTEDTARTANAL	951
D	b	887	GTGICYOVPISGFSFDMERNHRPDELYLTERPAANMEANKPCTPLTITTEDVARTANAL	946
Q	y	952	EIDATEVGRACAGCTISPGIYHOTOFTAGYPGSGKRSITQCGDVVVYPPRTRELSNMR	1011
D	b	947	ELDSATVGRACACRYTPGVYOFTAGYPGSGKRSITQADVDDVVYPPRTRELSNMR	1006
Q	y	1012	RGFAFPTHPTAARTITRRVYIDEASLPHLLHLMQASSVHLLGDPNOIPIADEFHA	1071
D	b	1007	RGFAFPTHPTAARTITGRRAVYIDEASLPHLLHLMQAAATVHLLGDPNOIPIADEFHA	1066
Q	y	1072	GLVPAIRPELAPTSMWAVVYHRCRPADVCELRGAYPRKIQTSSVYLSLEFNPEPAIOCKLVX	1131
D	b	1067	GLVPAIRPELAPTSMMHVTYHRCRPADVCELRGAYPRKIQTSSVYLSLEFNPEPAIOCKLVX	1126
Q	y	1132	TOAKKANPGATVYHEAQATFTEETIITATADAKGLIOSSRAHAIVALTREHEKCVILDA	1191
D	b	1127	TOAKKANPGSVYVHEAQATFTEETIITATADAKGLIOSSRAHAIVALTREHEKCVILDA	1186
Q	y	1192	PGLLREGISDVIYVNNFFLAGGVGWHRSVYPRGPDONLCTIDAFPRSCQISAYHQLA	1251
D	b	1187	PGLLREGISDVIYVNNFFLAGGVGWHRSVYPRGPDONLCTIDAFPRSCQISAYHQLA	1246
Q	y	1252	EELCHRPAPAAVLPPELEQGLKMPQELVYSDSVLEFELTDIVHGMAPSORKAVL	1311
D	b	1247	EELCHRPAPAAVLPPELEQGLKMPQELVYSDSVLEFELTDIVHGMAPSORKAVL	1306
Q	y	1312	STLVGRYGRKLYLEAASDVRESLARFTPTIGPVQATTCEDLYELVEAMVEKGODGSAYL	1371
D	b	1307	STLVGRYGRKLYLEAASDVRESLARFTPTIGPVQATTCEDLYELVEAMVEKGODGSAYL	1366
Q	y	1372	ELDLCNDVSRITFPQKXCKFTTGETITAHGVGOCISAMSKTFCALFGPMFRALEKITL	1431
D	b	1367	ELDLCNDVSRITFPQKCKFTTGETITAHGVGOCISAMSKTFCALFGPMFRALEKITL	1426

QY	1432	ALLPENIYEGDAEESYFAAAYSGACSSCFENDSEFPSTONNNSLIGECVMECGM	1492
		: : : : : : : : :	
Db	1427	ALLPGCYFYEGDAEDDYFSAVAANAARSMYFENDSEFPSTONNNSLIGECIMVCGMP	1466
QY	1492	QMLIRLYHLVRSAMITLOAPKESLKGFWKXHGSEPGTLLMNTVWMAAIIAHCYEFPDFRYA	1551
Db	1487	QMLIRLYHLIRSAMITLOAPKESLRGFWKXHGSEPGTLLMNTVWMAAVITHCYEDFDLOVA	1546
QY	1552	AFKDDSDVILCSODVROSRNAAALIACCGIKLKVDYRPTLYAGVVAAGLGTLPVRYFA	1611
Db	1547	AFKDDSDVILCSSEYRQSPGAAVILIACCGIKLKVDYRPTLYAGVVAABGLALPVRVFA	1606
QY	1612	GRISEKMWGPEPERAEQJLRLAVCDFTLGLTTNAQVCVDVVSRYGVSPGLVHNLIGMLQT	1671
Db	1607	GRLETKMWGPEPERAKOLRLAVSDFLRKLTTNAQOMCVDVVSRYGVSPGLVHNLIGMLQA	1666
QY	1672	IADGKAHFEETIKRVIDLNTSIIOARE	1698
Db	1667	VADGKAHFTESVAKRVIDLNTSILCRVE	1693

RESULT	13		
069418			
ID	069418	PRELIMINARY:	PRT: 1693 AA.
AC	069418;		
DT	01-NOV-1996 (TREMBLrel. 01, Created)		
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)		
DT	01-JUN-2000 (TREMBLrel. 14, Last annotation update)		
DE	ORF1, ORF2 & ORF3.		
OS	Hepatitis E virus.		
OC	Viruses; ssRNA positive-strand viruses, no DNA stage.		
OX	NCBI_TaxID=12461;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=INDIVIDUAL PATIENT INFECTED WITH HEV.		
RA	Von Bunn A., Seebach J., Thyagarajan S.P., Mohanavalli B., Menon T.,		
RA	Froesner G.;		
RL	Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.		
DR	EMBL: X99441; CAA67802.1, -		
DR	INTERPRO: IPR000606; -		
DR	INTERPRO: IPR002388; -		
DR	INTERPRO: IPR002589; -		
DR	PFAM: PF01443; Viral_helicase1; 1.		
DR	PFAM: PF01660; Usmethyltransf; 1.		
DR	PFAM: PF01661, DUF27, 1.		
QO	SEQUENCE 1693 AA; 18549 MW; 489B4ADB5E7E529 CRC64;		

	Query Match	81.0%;	Score 7301;	DB 14;	Length 1693;	
	Best Local Similarity	79.9%;	Pred. No. 0;	Mismatches 188;	Indels 40;	Gaps 6;
	Matches 1367;	Conservative 116;				
QY	1 PGITTAIEQAALAAANSALANAVVPRPELSRQTEIILINLMOPROLVREPLYMNPRIOR					
Dd	10 PGITTVIEQAALAAANSVLANNAVVRPLSHOITEILINLMOPROLVRRPEVFMNHPRIOR					
QY	61 VIHNELEOYCRAAGRCLEVGAHPRSINDNPVNLHRCFLRPVGRDVQRYSAPTRGPAN					
Dd	70 VIHNELELYCARASRCLEIGHAPRSINDNPVVHARCFLRPVGRDVGMRHNAPTRGPDAN					
QY	121 CRRSALNGLPADRTYCFDFGFSRCFAAETGVALLYSLHDMLPADVAEAMRHGXTRLTYAA					
Dd	130 CRRSALRGLPADRTYCFDFGFGSCNFPAETGVALLYSLHDMSPSDVAEAMFRHGCMRLTYAA					
QY	181 LHLPPVLLIPPGYTHHTSYLLIHGDRAVVITYEGGTSGAGYNHDSILEAMIRTKTIIVGDH					
Dd	190 LHLPPVLLIPGTYRTASYLLIHGDRVVVITYEGGTSGAGYNHDSVNLMSWIKTTKTVTDGH					
QY	241 PLVIERYVAICHCYVLLTLTAAPEPSPMYVYPRSTEVYYVRSIFEPGGSPSLFSPCASCTK					
Dd	250 PLVIERYKAICHCVLTLTLTAAPEPSPMYVVPPIPNSTEVYVRSIFEPGGSPSLFSPCASCTK					

QY	301	STFAVNVHIMLDRIMLEGAATLDQACRSRMITYLRGIS KYTVGATLVANBEMANSEAL	360
Db	310	STFAVNAHIMDRIMLEGAATLDQACRSRMITYLRGIS KYTVGATLVANBEMANSEAL	369
QY	361	TAKTTAAVLTICHOYRLTQALISKGRRLGVEHAOKFITRLTSMLEFGSGRDYIPGRLO	420
Db	370	TAVITAAVLTICHOYRLTQALISKGRRLERHAOKFITRLTSMLEFGSGRDYIPGRLO	429
QY	421	FYAOCRMLISAGHILDRVLVEDESVPCORITFLKKVAGFCCFMRMLGOECTCELEPAE	480
Db	430	FVYOCRRMLISDGHILDRVLVEDESPACRCRIVIRKALSCFCFMRMLGOECTCELEPAE	489
QY	481	GLVGDHNDHNEAVESSEVDPAPAHILNDVSGTAVHGHOLEALRYLNLNPDIAANASLT	540
Db	490	GVGDGDHNDHNEAEGSDVDPAPAESALISDGSYVSGTDLPRLYQLIDRDELVAACRLI	549
QY	541	ATVELVASPRLCECRIVLGKRTFRTVYDGAHLEANGPEOYALSPDASROSGASHLT	600
Db	550	ATVAVSVDORICOCETILGKTRTGFVGVAVLETNNGPRHMLSPDASOSTYAGSLST	609
QY	601	YELTPAGLOVRISNSNGDCTATFPPGAGAPSAAPGEVAAFCSALYRYNRETOHNSLTGLM	660
Db	610	YAVSADLEVRVYAAAGDHNRAVAPASVSPRSPGVEYATCASCALYRFRNREVOHNSLTGLM	669
QY	661	LHPRGLGITPPSPSGHIMSANPCGECTYITRTMS -TSGSSDSP-----PEANA	714
Db	670	FHRPGLITGLPAPSPGVWBSANPCFGEESTYITRTMSVEDAVASSPARDLGVSBSIPS	729
QY	715	MAAPRGLHSTPSPVSDIWLPRPSEFEOYDAAPV -PRAPDPAGLGCVLTPRPPRPNK	773
Db	730	RAAPRPL-----AAPLRPADSPRPPS-----APADPRPSG	762
QY	774	PSLPP-----SRNRLLTYTPDGAKYVAGSLFESDCMLVNASNPGHRRPGGSLCHAFY	827
Db	763	TTAGAPALITHQMARHRRLLFTYPDGSKVAGSLFESTCTWPNASNVDBRRPGGSLCHAFY	822
QY	828	QRPEAPATPREFITMRBGLAATITLTPRPIIHANAPOYRVQONKRLDEAATRETSKRGTA	887
Db	823	QRYPASDASAFWRGCAAAATLTPRPIIHANAPDRLRHNKRLDEAATRETSKRLGTA	882
QY	948	NLALEIDAATEVGRACACTISPGLVHYOFTAGVPGSGSRSIOGDVDVVVPTRELRN	1007
Db	943	NLALEIDASIDVGRACAGCVTPGVYQOFTAGVPGSGSRSITQADVDVVVPTRELRN	1002
QY	1008	SWRRRGCAFTPTHTAARVITGRNVIDEAPSLRPHLLHMRASSVHLLGDPNOIPAD	1067
Db	1003	AMRRRGCAFTPTHTAARVITGRNVIDEAPSLRPHLLHMRASSVHLLGDPNOIPAD	1062
QY	1068	FEHAGVLPAIRPELAPTSMMXVYHRCRPOVCELTIRAYK IOTGSRVLYRSLFWNPALGO	1127
Db	1063	FEHAGVLPAIRPDLAPTSMMXVYHRCRPOVCELTIRAYMIDITSRVLYRSLFWNPALGO	1122
QY	1128	KLVYTOAKAANPGATITVEHAOGATTEETITATADARLEIOSSRAHAIVALTRHTEKV	1187
Db	1123	KLVYTOAKAANPGSVYVEHAOGATTEETITATADARLEIOSSRAHAIVALTRHTEKV	1182
QY	1188	ILDAARGLREVGISDYIVNNFPLAGEVGHRRPSVTPRGNDONITGLQAPPSQISAY	1247
Db	1183	ILDAARGLREVGISDYIVNNFPLAGEVGHRRPSVTPRGNDONITGLQAPPSQISAY	1242
QY	1248	HQLEELGHRPAPVAANLPCRCPLBEGLLYMOPELTVSDSVLYEFLTITVHCRMAAPSR	1307
Db	1243	HQLEELGHRPAPVAANLPCRCPLBEGLLYMOPELTVSDSVLYEFLTITVHCRMAAPSR	1302
QY	1308	KAVYSLTVGRYGRKTKYLEAHASDRESLARFPTIGPQVATTCLELYELVEMVKGODG	1367
Db	1303	KAVYSLTVGRYGRKTKYINSHSDRESLARIRIIPAGPQVYTTCLELYELVEMVKGODG	1362
QY	1368	SAVLELDLCNBDVSRITFEOKXCNKFTTGEBITAHCKVGOISAMSKTCALGFWPRAIE	1427

```

Db   1363 SAVLELHMNDVSTITFFQDDCNKFTTGEIAHAKKVQGSIASNKTCPLEGPFRAIE      1422
OY   1428 KEIALLPNNFYGAYEESVFAAVASGCSCGMFENDFSEFDSTONNFSLGLECVMAE     1487
       ||||| :|||||:::||| | :|||||:::||| | :|||||:::||| | :|||||:::||| |
Dd   1423 KAILLLDPGVGYFGDAFPDYTFSAAVPAKASMVEFNDFSSEDSTQNNSFLSGLCAIMEE    1482
OY   1488 CGMPQMILRLYLHLVSAMIILOAPKESELKGPMKKHSSEPDTLLMNTVMNAIIAHCEFRD    1547
       ||||| :|||||:::||| | :|||||:::||| | :|||||:::||| | :|||||:::||| |
Dd   1483 CGMPQMILRLYLHLRSAMILOAPEKSLETGFMKKHSGSEGPTLLMNTVMNAVITHCYDEPD    1542
OY   1548 FRVAAFKCDSBVCISDVROSRNNAALLAGCINKVDYRPIIGLYAGVVAPRGJLTEDY      1607
       ||||| :|||||:::||| | :|||||:::||| | :|||||:::||| | :|||||:::||| |
Dd   1543 LQVAAPFKSDSVICSEFSROSPPGAIVLLAAGCIKLTKPDFRPtGIYGAVVPARGALPDV     1602
OY   1608 VRFGRARLKSEMNGCFPEAEDELTLAVCDFELGTINFAOVCDVYSRVYGSPLGHNLIG     1667
       ||||| :|||||:::||| | :|||||:::||| | :|||||:::||| | :|||||:::||| |
Dd   1603 VRFAGRLEKRWMGCGPERAEDELRLAVSPFLKLNIVAOMCDVYSRYVGSPGLYHNLI G     1662
OY   1668 MLQTADGKAHFETIKTPVLDLTNISIORVE 1698
Dd   1663 MLDVAADGKAHFETSVKRPVLDTNSILCRVE 1693
RESULT 14
ID ID          PRELIMINARY;           PRT;         727 AA.
AC O81873;
OT O81873;
DT 01-NOV-1996 (TREMBREL .01, Created)
DT 01-NOV-1996 (TREMBREL .01, Last sequence update)
DT 01-JUN-2000 (TREMBREL .14, Last annotation update)
DE POLYPEPTIDE
DE (ENCODING NUP-BINDING PROTEIN AND RNA-DEPENDENT RNA POLYMERASE)
DE (FRAGMENT).
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
CX NCBI_Taxid=12461;
RN [1]
RP SEQUENCE FROM N.A.
RA Tam A.W., Smith M.M., Kim J.P., Young L.M., Platak M., Feldman R.A.,
RA Purdy M.A., Bradley D.W., Reyes G.R., Fry K.E.;
RU Proc. Natl. Acad. Sci. U.S.A. 0:0-0(0).
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90193694; PubMed=2107574;
RA Reyes G.R., Purdy M.A., Kim J.P., Luk K.C., Young L.M., Fry K.E.,
RA Bradley D.;
RT "Isolation of a cDNA from the virus responsible for enterically
RT transmitted non-A, non-B hepatitis." ;
RL Science 247:1335-1339(1990).
[3]
RP SEQUENCE FROM N.A.
RX MEDLINE=92271462; Pubmed=1589964;
RA Fry K.E., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,
RA Platak M., Feldman R.A., Yun K.Y., Purdy M.A., Et AL.;
RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
RT region encoding consensus motifs for an RNA-dependent RNA polymerase
RT and an ATP/GNP binding site.";
RL Virus Genes 6:173-185(1992);
DR EMBL; M32400; AAA03206.1; -.
DR INTERPRO; IPR000606; -.
DR PFAM; PF01443; Viral_helicasel_1.
KW Polyprotein; RNA-directed RNA polymerase.
FT NON_TER                      1
SO SEQUENCE              727 AA; 79306 MW; 872691f6c2318fa7 CRC64;
```

```

Db 1 VVOYQFTAGVPGSGKSSITQADVDVVVPTRELINAMRRRCFAAFTPHPAARTVTOGRV 60
QY 1032 VIDEAPSLPRLHLLHMQARASVHLLDPPQIADIDEHAGLVPAIRPELAPTSMMXVTH 1091
Db 61 VIDEAPSLPRLHLLHMQARATVHLLDPPQIADIDEHAGLVPAIRPDGPTSMWVTH 120
QY 1092 RCPADVCCELLRGAPKIQTTSRVLSLFENEPALGKLVYTOAKKANPCALTYHEAOGA 1151
Db 121 RWPADVCCELLRGAPMIQTTSRVLSLFENEPALGKLVYTOAKKANPCALTYHEAOGA 180
QY 1152 TPEETTTIATADAGLQSSRAHVAIVLTHTEKCVILDPGLIREGISDVVNNFELA 1211
Db 181 TYETTTIATADAGLQSSRAHVAIVLTHTEKCVILDPGLIREGISDVVNNFELA 240
QY 1212 GGEVGHKRPVIRPGNDQNLGTLOAPPCQISAVYHQAELGHRPAPAAVLPCCPEL 1271
Db 241 GGEVGHKRPVIRPGNDQNLGTLOAPPCQISAVYHQAELGHRPAPAAVLPCCPEL 300
QY 1272 EOGLLYMPQELTVSDSVLVELTDIVHCRMAAPSORRAVLSLVGRYGRRTKLYEAASD 1331
Db 301 EOGLLYMPQELTVSDSVLVELTDIVHCRMAAPSORRAVLSLVGRYGRRTKLYEAASD 360
QY 1332 VRSLARFPTTGPVQATTELELYELVAMYEKGDGSAVLELDCNDVSRITPFQKXCN 1391
Db 361 VRSLARFPTTGPVQATTELELYELVAMYEKGDGSAVLELDCNDVSRITPFQKXCN 420
QY 1392 KFTTGTETIAGKYGOGISAMSKTFCAFGPMFRAIEKEIITALLPNTFYGDVAYEESVFAA 1451
Db 421 KFTTGTETIAGKYGOGISAMSKTFCAFGPMFRAIEKEIITALLPNTFYGDVAYEESVFAA 480
QY 1452 AVSAGSCWYFENDESEFSDSTQNNFSLGLECYVMEECGMQWILRLVLSRANIIQAPK 1511
Db 481 AVSAGSCWYFENDESEFSDSTQNNFSLGLECYVMEECGMQWILRLVLSRANIIQAPK 540
QY 1512 ESKLGFKKKSGEGTLLMTVMMAIIAHCYERDRFRAVAFGDDSVILCSOVRORNA 1571
Db 541 ESKLGFKKKSGEGTLLMTVMMAIIAHCYERDRFRAVAFGDDSVILCSOVRORNA 600
QY 1572 AALIAGCGLKLVDRPFGIYAGVAVPAGLTLDPVVRFAFGRLESEKMWGSGPERAEOLRL 1631
Db 601 AALIAGCGLKLVDRPFGIYAGVAVPAGLTLDPVVRFAFGRLESEKMWGSGPERAEOLRL 660
QY 1632 ACVDFLGLTNVAVCVVYVSRVYGVSPGLVHNLIGMLQTLADGKAHFTETIKVVDLTN 1691
Db 661 ACVDFLGLTNVAVCVVYVSRVYGVSPGLVHNLIGMLQTLADGKAHFTETIKVVDLTN 720
QY 1692 SIIOQVE 1698
Db 721 SIIOQVE 727

```

```

RESULT 15
071144 PRELIMINARY: PRT: 479 AA.
AC 071144,
DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DE 01-MAY-2000 (TREMblrel. 13, last annotation update)
DE NONSTRUCTURAL POLYPROTEIN (FRAGMENT).
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
ON NCBI_TaxID=12461;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=US-1.
RX MEDLINE=98178637; PubMed=9519822;
RA Schlauder G.G., Dawson G.J., Erker J.C., Kwo P.Y., Knigge M.F.,
RA Smally D.L., Rosenblatt J.E., Desai S.M., Mushahwar I.K.;
RT "The sequence and phylogenetic analysis of a novel hepatitis E virus
RT isolated from a patient with acute hepatitis reported in the United
RT States";
RL J. Gen. Virol. 79:447-456(1998).
DR EMBL; AF035436; AAC40598.1; -

```

```

DR INTERPRO: IPR002588;
DR PFAM: PF01660; Vmethyltransf. 1.
KW Polypeptidase.
FT NON_TER 1
FT NON_TER 479
SQ SEQUENCE 479 AA: 53778 MW: 2180048A326037A4 CRC64:

```

```

Query Match 28.0%; Score 2520; DB 14; Length 479;
Best Local Similarity 99.8%; Pred. No. 5, 2e-178;
Matches 471; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 PGTTTAEQALAAANSALANAVVPRFLSRVOTELINIMQROLYFREVMNPHIOR 60
Db 8 PGTTTAEQALAAANSALANAVVPRFLSRVOTELINIMQROLYFREVMNPHIOR 67
QY 61 VINNELQYCARARAGRCLEVGAPRSTINDPNVLRHCELPVRDVOQWRSAPTRGPAN 120
Db 68 VINNELQYCARARAGRCLEVGAPRSTINDPNVLRHCELPVRDVOQWRSAPTRGPAN 127
QY 121 CRSSALRGLPADRTYCFDGSRCAPFAETGVALYSLHDLMPADVAEAMARHGXTRLYAA 180
Db 128 CRSSALRGLPADRTYCFDGSRCAPFAETGVALYSLHDLMPADVAEAMARHGXTRLYAA 187
QY 181 LHPPEVLLPPGTYYHTTSYLLIHDGRAVVTYEGDTSAGYNHDVSLIRANIRTTKIVGDH 240
Db 188 LHPPEVLLPPGTYYHTTSYLLIHDGRAVVTYEGDTSAGYNHDVSLIRANIRTTKIVGDH 247
QY 241 PLYTERRAIGCHFVLLTLTAAPSPMPYVPRSTEVYRSTFEGPGSPLEPSCSTK 300
Db 248 PLYTERRAIGCHFVLLTLTAAPSPMPYVPRSTEVYRSTFEGPGSPLEPSCSTK 307
QY 301 STEHAAPVHIWDRMLFEGATLDQAFCCSRMLTYLRGISYKVTIVGALVANEGNNASEDAL 360
Db 308 STEHAAPVHIWDRMLFEGATLDQAFCCSRMLTYLRGISYKVTIVGALVANEGNNASEDAL 367
QY 361 TAXITAAVLTICHQRYLRTQALSKGMRLEVEHNAOKFTFLYLSMLEKSGRDYIPGROLQ 420
Db 368 TAXITAAVLTICHQRYLRTQALSKGMRLEVEHNAOKFTFLYLSMLEKSGRDYIPGROLQ 427
QY 421 FYVQCRRLMSAGFHLDPRLVLPDESVPCCRTFLKKVAGKFCCEMRLAGQEC 472
Db 428 FYVQCRRLMSAGFHLDPRLVLPDESVPCCRTFLKKVAGKFCCEMRLAGQEC 479

```

Search completed: May 30, 2001, 16:14:01
Job time: 278 sec